

Yongheng Wang  
Xiaoming Zhang (Eds.)

Communications in Computer and Information Science

312

# Internet of Things

International Workshop, IOT 2012  
Changsha, China, August 2012  
Proceedings



Springer

Communications  
in Computer and Information Science 312

Yongheng Wang Xiaoming Zhang (Eds.)

# Internet of Things

International Workshop, IOT 2012  
Changsha, China, August 17-19, 2012  
Proceedings

Volume Editors

Yongheng Wang  
Hunan University  
College of Information Science and Engineering  
Lushan Road Changsha  
410082 Changsha, China  
E-mail: yh.wang.cn@gmail.com

Xiaoming Zhang  
Hunan University  
College of Information Science and Engineering  
Lushan Road Changsha  
410082 Changsha, China  
E-mail: zhangxm19712003@yahoo.com.cn

ISSN 1865-0929  
ISBN 978-3-642-32426-0  
DOI 10.1007/978-3-642-32427-7  
Springer Heidelberg Dordrecht London New York

e-ISSN 1865-0937  
e-ISBN 978-3-642-32427-7

Library of Congress Control Number: 2012943580

CR Subject Classification (1998): C.2, C.3, D.4.6, J.1, H.3-5

© Springer-Verlag Berlin Heidelberg 2012

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, 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.

*Typesetting:* Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

# Preface

The Internet of Things (IoT) is a novel paradigm that aims at bridging the gap between the physical world and its representation within the digital world. The idea is to integrate the state of the “things” that form the world into software applications, making them benefit from the world’s context information. IoT technology is developing rapidly and many applications have been built recently. But there are still many challenges in the fields of wireless sensor networks (WSN), radio-frequency identification (RFID), data processing, security, etc. The 2012 International Workshop on Internet of Things (IOT Workshop 2012) brought together researchers and industry pioneers to discuss this important frontier.

The present volume contains the papers that were independently peer-reviewed and selected by the IoT workshop 2012 Program Committees. Forty-one percent of paper submissions were accepted.

As Program Co-chairs we would like to thank the members of the Program Committee for the hard work. We also thank all the authors who submitted papers to this workshop.

August 2012

Yongheng Wang  
Xiaoming Zhang

# **Organization**

The 2012 International Workshop on Internet of Things (IOT Workshop 2012) was organized by Hunan University, Post and Telecommunication, Beijing Normal University, Computer Engineering and Science Journal Press and Hunan Computer Federation.

## **Honorary Chairs**

Jianquan Yao	Fellow of Chinese Academy of Sciences, Tianjin University, China
Jianpin Ge	Vice President, Beijing Normal University, China

## **Conference Chairs**

Edwin Sha	University of Texas-Dallas, USA
Renfa Li	Hunan University, China
Mingquan Zhou	Beijing Normal University, China

## **Conference Steering Committee**

Xinxia Wang	Chinese Institute of Electronics, China
Dingyi Fang	Northwest University, China
Yanxiang He	Wuhan University, China
Jian Kuang	Beijing University of Posts and Telecommunications, China
Kejin Bao	Jiangsu University, China
Hongjun Wang	Shandong University, China
Guosheng Yang	Minzu University of China, China
Xiaoping Zeng	Chongqing University, China
Jiawei Luo	Hunan University, China
Huan Zhao	Hunan University, China

## **Program Committee Chairs**

Hai Jin	Huazhong University of Science and Technology, China
Jianguo Ma	Tianjin University, China
Ping Zong	Nanjing University of Posts and Telecommunications, China

Kenli Li Hunan University, China

## VIII Organization

Zhufeng Li  
Xiaoming Zhang  
Yongheng Wang

Beijing Normal University, China  
Hunan University, China  
Hunan University, China

## Program Committee

Shining Li  
Zhihong Qian  
Xiaolin Gui  
Jiali Bian

Northwestern Polytechnic University, China  
Jilin University, China  
Xi'an Jiaotong University, China  
Beijing University of Posts and  
Telecommunications, China

Zhiliang Wang

University of Science & Technology Beijing,  
China

Xiaohan Sun  
Cheng Xu  
Zhiwei Kang  
Shuguang Peng  
Abdoul Rjoub

Southeast University, China  
Hunan University, China  
Hunan University, China  
Hunan University, China  
Jordan University of Science and Technology,  
Jordan

Abdul Raouf Khan  
Ahmad Fairuz Bin Omar  
Ahmed Mansour Manasrah  
Bilal Bahaa Zidan  
Bindu Garg  
Chia-Sheng Tsai  
Chunliang Wu  
D.P. Sharma  
D. Balaji  
Jude Hemanth  
Huanjing Wang  
Irving Vitra Paputungan

King Faisal University, Al Hasa, Saudi Arabia  
University of Science and Technology, Malaysia  
Botnet Research, Malaysia  
University of Malaya, Malaysia  
ITM Engineering College, New Delhi, India  
Tatung University, Taiwan  
Guangdong Ocean University, China  
ASBM, Hyderabad, India  
College of Engineering and Technology, India  
Karunya University, India  
Western Kentucky University, USA  
Islamic University of Indonesia, Yogyakarta,  
Indonesia

J. Hasan  
Kilari Veeraswamy

University of Economics in Bratislava, Slovakia  
QIS College of Engineering and Technology,  
India

Lim Tek Yong  
Nagendra Sah  
Saroj Kumar Gupta  
Seyed Amin Hosseini Seno  
Shah Ahsanal Haque

Multimedia University, Malaysia  
PEC-Deemed University, India  
MITS, Madanapalle, AP, India  
Universiti Sains Malaysia, Malaysia  
International Islamic University Chittagong,  
Bangladesh

Vilas Lambture  
Virendra P. Vishwakarma  
Wafa Barkhoda

Sir Visvesvaraya Memorial Engineering College,  
Nashik, India  
Guru Gobind Singh Indraprastha University,  
India  
University of Kurdistan, Iran

# Table of Contents

## Wireless Sensor Networks

Wireless Sensing and Propagation Characterization for Smart Greenhouses.....	1
<i>Yuyang Peng, Ushan He, and Jaeho Choi</i>	
Efficient Ad Hoc Routing Technique for Connecting Geographically Distributed Heterogeneous Networks .....	9
<i>Youngshin Ahn, Yi Zhao, Ushan He, and Jaeho Choi</i>	
An Adaptive Scheduling Algorithm for the Patient Monitoring System on WBANs .....	17
<i>Hongkyu Jeong</i>	
An Efficient Data Aggregation Scheme in Wireless Sensor Networks ....	25
<i>Ying Wang and Guorui Li</i>	
Design of the Remote Monitoring System for Workshop Based on ZigBee Wireless Sensor Networks .....	33
<i>Hongcheng Yu, Haiping Zhu, Fei He, and Yunlong Wan</i>	
Delay Aware Time Slot Allocation Algorithm in Light-Trail Network ...	41
<i>Minglei Fu, Yiluan Zhuang, Maolin He, and Zichun Le</i>	
Quality of Recovery (QoR) Analysis for the ONU Protection in WOBAN .....	49
<i>Zichun Le, Maolin He, Yiluan Zhuang, and Minglei Fu</i>	
Energy Efficient Hop Length Optimization for Laterally Connected Wireless Sensor Networks .....	57
<i>Yuyang Peng, Ushan He, and Jaeho Choi</i>	
Performance Evaluation Analysis about Ethernet and DeviceNet .....	64
<i>Wen Li and Xiangyu Dai</i>	
Network Interaction Analysis of 3G and WiFi .....	70
<i>Xuejun Meng</i>	
Ant Intelligence Routing Algorithm for Wireless Sensor Networks .....	76
<i>Awudu Karim, Xiaoming Zhang, A.M. Oluyemi, and T. Fitarikandro</i>	
Research on Difference Distance Measurement and Localization Based on Zigbee .....	86
<i>Zeng Gang</i>	

Routing Optimization Based on Ant Colony Algorithm for Wireless Sensor Networks with Long-Chain Structure .....	91
<i>Jing Gao, Limin Wei, Yongli Zhu, and Lifen Li</i>	
A Key Management of Wireless Sensor Networks for Telemedicine Care .....	98
<i>Min Nan, Yong Xu, and Pengcheng Zhao</i>	
An Improved Routing Algorithm on LEACH by Combining Node Degree and Residual Energy for WSNs .....	104
<i>Weiping Luan, Changhua Zhu, Bo Su, and Changxing Pei</i>	
Priority-Based Random Access Algorithm for TD-SCDMA Trunking System .....	110
<i>Qing Jiang, Xianglin Wu, Hao Chen, and Xueqian Wang</i>	
Research on Improved DV-HOP Localization Algorithm Based on the Ratio of Distances .....	118
<i>Yan Hu, Zhilong Shan, and Hua Yu</i>	
Survivability Evaluation of Cluster-Based Wireless Sensor Network under DoS Attack .....	126
<i>Chunjie Chang, Changhua Zhu, Honggang Wang, and Changxing Pei</i>	
Research on Ubiquitous Network Technique and Application .....	133
<i>Ping Zhu and Xiaofei Xu</i>	
Design and Implementation of the Patrol Terminal System for Power Transformation Facilities Based on Android and Wireless Network .....	141
<i>Jiantao Zhao, Lin Bian, and Yue Lian</i>	
Dynamic One-Way Key Establishment Scheme in Wireless Sensor Networks .....	149
<i>Sisi Jiang, Zhengye Si, Zushun Wu, and Dan Li</i>	
Query-Aware Location Privacy Model Based on p-Sensitive and k-Anonymity for Road Networks .....	157
<i>Jiahui Chen, Hongyun Xu, and Lin Zhu</i>	
A Wireless Measurement System (M3D) for Three-Dimensional Gait Analysis System .....	166
<i>Tao Liu, Yoshio Inoue, Kyoko Shibata, and Kouzou Shiojima</i>	
Approaches and Controllers to Solving the Contention Problem for Packet Switching Networks: A Survey .....	172
<i>Fouad Kharroubi, Lin Chen, and Jianjun Yu</i>	

Experimental Analysis of AODV, DSR and DSDV Protocols Based on Wireless Body Area Network .....	183
<i>Clement Ogugua Asogwa, Xiaoming Zhang, Degui Xiao, and Ahmed Hamed</i>	
OCTBR: Optimized Clustering Tree Based Routing Protocol for Wireless Sensor Networks .....	192
<i>Jian Zhang, Yu Xie, Dandan Liu, and Zhen Zhang</i>	
An Intelligent Irrigation System for Greenhouse Jonquil Based on ZigBee Wireless Sensor Networks .....	200
<i>Zongyu Xu, Baodong Lou, and Guangcheng Shao</i>	
M/I Adaptation Layer Network Protocol for IoT Based on 6LoWPAN .....	208
<i>Zhihong Qian, Yijun Wang, Xue Wang, and Shuang Zhu</i>	

## RFID

A Novel Method to Improve Gain and Tune Impedance of RFID Tag Patch Antenna .....	216
<i>Chuncheng Kong and Jun Hu</i>	
Improved Accuracy of RFID Localization Assisted by Output Power Adjustment of the Reader .....	220
<i>Xiaoyin Li, Lianshan Yan, Wei Pan, Bin Luo, and Q.F. Guo</i>	
A Two-Layer Duplicate Filtering Approach for RFID Data Streams .....	226
<i>Wen Jiang, Yongli Wang, and Gongxuan Zhang</i>	
RFID Uncertain Data Cleaning Framework Based on Selection Mechanism .....	234
<i>Xiufeng Xia, Lijuan Xuan, Xiaoming Li, and Ying Li</i>	
An Effective Temporary ID Based Query Tree Algorithm for RFID Tag Anti-collision .....	242
<i>Yun Tian, Gongliang Chen, and Jianhua Li</i>	
An Anti-collision Algorithm of RFID Tags Based on CDMA .....	248
<i>Weijun Zhang, Shuping Zhang, and Dawei Zhang</i>	
The Research on Electronic Tag Quantity Estimate Arithmetic Based on Probability Statistics .....	254
<i>Lin Zhou, Zhen Li, Yingmei Chen, and Tong Li</i>	
An Improved RFID Data Cleaning Algorithm Based on Sliding Window .....	262
<i>Lingjuan Li, Tao Liu, Xiang Rong, Jianxin Chen, and Xiaolong Xu</i>	

Methods to Recognize Special Tags in UHF RFID System . . . . .	269
<i>Lei Hu, Zhen Huang, and Bowen Chen</i>	

## Sensors and Equipments

Sensor Ontology Building in Semantic Sensor Web . . . . .	277
<i>Yimin Shi, Guanyu Li, Xiaoping Zhou, and Xianzhong Zhang</i>	
Panoramic CIM Model of Power Equipment at Converter Station Based on IOT . . . . .	285
<i>Jingyu Huang, Chun Huang, Xiaoqing Huang, Junyong Zhang, and Jie He</i>	
An Infrared Ranging System for Automotive Anti-collision . . . . .	293
<i>Liang Xu and Zhiqiang Meng</i>	
Fault Diagnosis for Power Equipment Based on IoT . . . . .	298
<i>Yusheng Zhu, Xiaoqing Huang, Junyong Zhang, Jie Luo, and Jie He</i>	

## Data Processing

Fast QR Code Image Process and Detection . . . . .	305
<i>Qichao Chen, Yaowei Du, Risan Lin, and Yumin Tian</i>	
Towards Adaptable Workflow Management System: Shark Enhydra . . . . .	313
<i>Lazarus Obed Livingstone Banda, Zuping Zhang, and Jing Xia</i>	
Quantized Communication of Multi-agent Systems under Switching Topology . . . . .	321
<i>Qian Ye, Xuyang Lou, and Baotong Cui</i>	
Research on Automotive Parts Abradability on Driving Behavior Analysis . . . . .	327
<i>Hu Wang and Xuan Zhang</i>	
Existence and Stability of Equilibrium of Discrete-Time Neural Networks with Distributed Delays . . . . .	334
<i>Xuyang Lou, Baotong Cui, and Qian Ye</i>	
Construction Scheme and Key Technologies of Electric Energy Information Acquisition System . . . . .	340
<i>Enguo Zhu and Xuan Liu</i>	
An Effective Algorithm of Outlier Detection Based on Clustering . . . . .	346
<i>Qingsong Xia, Changzheng Xing, and Na Li</i>	
Research on Multimedia Signal Acquisition Strategy Based on Compressed Sensing . . . . .	352
<i>Xiaohua Guo</i>	

PFSA: A Novel Fish Swarm Algorithm .....	359
<i>Zushun Wu, Zhangji Zhao, Sisi Jiang, and Xuechi Zhang</i>	
Research on Platform Integration of Equipment Support Simulation Training Systems and Integration Degree .....	366
<i>Yunfeng Lian, Yu Lu, LiYun Chen, and Yi Ma</i>	
A New Method for Extraction of Signals with Higher Order Temporal Structure .....	372
<i>Lei Hu, Bowen Chen, and Zhen Huang</i>	
A New Method for Vibration Signal Analysis Using Time-Frequency Data Fusion Technique .....	380
<i>Lei Hu, Bowen Chen, and Zhen Huang</i>	

## Security

A Multi-layer Security Model for Internet of Things .....	388
<i>Xue Yang, Zhihua Li, Zhenmin Geng, and Haitao Zhang</i>	
Security Research on Cloud-Based Logistics Service Platform .....	394
<i>Fuquan Sun, Chao Liu, Xu Cheng, and Dawei Zhang</i>	
Security Technology Analysis of IOT .....	401
<i>SheQiang Peng and HongBing Shen</i>	
A Study on Mobile Phone Security Industrial Ecology .....	409
<i>Qiyu Chen, Jinlong Hu, and Ling Zhang</i>	
Research on Sensor-Gateway-Terminal Security Mechanism of Smart Home Based on IOT .....	415
<i>Fei Li, Zhou Wan, Xin Xiong, and Jiajun Tan</i>	

## Applications and Others

Indicator Framework of IOT Industry Growth Based on Analytic Hierarchy Process .....	423
<i>Jiajun Li and Nan Ma</i>	
Research on Architecture of the Internet of Things for Grain Monitoring in Storage .....	431
<i>Baisen Xu, Dexian Zhang, and Weidong Yang</i>	
Cloud-Recording Based Intelligent Feedback System of Voice Information .....	439
<i>Mengwei Si, Wenwen Du, Yibo Wu, and Jiawei Chen</i>	

Shopping System Based on Location Finding Technique of Internet of Things .....	445
<i>Bin Wu, Xiao Ling, HongWei Jia, and QiJin Sun</i>	
Study on the Relationship between Economic Growth and Carbon Emissions Based on Cointegration Theory .....	453
<i>Wenzhou Yan and Yiqing Deng</i>	
The Application of Cloud Computing in Smart Grid Status Monitoring .....	460
<i>Hongwei Bai, Zhiwei Ma, and Yongli Zhu</i>	
Design Smart City Based on 3S, Internet of Things, Grid Computing and Cloud Computing Technology .....	466
<i>Min Hu and Chang Li</i>	
Analysis and Design of Personalized Recommender System Based on Collaborative Filtering .....	473
<i>Jiantao Zhao, Hengwei Zhang, and Yue Lian</i>	
The Research of Data Mining Technology of Privacy Preserving in Sharing Platform of Internet of Things .....	481
<i>Luyu Chen and Guangwei Ren</i>	
The SVM and Layered Intrusion Detection System Based on Network Hierarchical .....	486
<i>Chao Ju Hu and Jin Wang</i>	
Application of Multi-vehicle Problem with Different Capacities .....	494
<i>Fuxing Yang and Bowen Yu</i>	
Study on Risk Control of Network Transactions Based on Customer Perspectives .....	498
<i>Mengting Sun</i>	
FPGA-Based Design and Implementation of Video Wall Display .....	503
<i>Yang Li and Xuesen Cai</i>	
Research on B-tree in Embedded Database SQLite .....	509
<i>Di Nan</i>	
The Method of Parallel Design Based on Simulation .....	516
<i>Jingmei Li, Qi Zhang, and Nan Di</i>	
Research on the Object-Oriented Unit Testing Based on the Genetic Algorithm .....	523
<i>Kehong Zhang</i>	
Data Aggregation and Information Type in Road Probing .....	529
<i>Lin Sun, Ying Wu, Jingdong Xu, Jinchao Li, and Yuwei Xu</i>	

The Smart Card Remote Unlocking Method and Its Implementation . . . . .	535
<i>Yongtao Hu, Xing Wang, and Yunlu Gao</i>	
A Rapid Review Method to Apply Digital Certificate Based on CA-Trust Set . . . . .	541
<i>Jingjing Yao, Yongtao Hu, and Yunlu Gao</i>	
A Navigation System on the Dynamic Constraint Condition . . . . .	547
<i>Qiang Ma, Jian Deng, and Gan Wei</i>	
How to Use Oprofile to Improve an Algorithm for Skia in Android . . . . .	553
<i>Haihang Yu and Gang Du</i>	
Cloud-Based Service Composition Architecture for Internet of Things . . . . .	559
<i>Li Liu, Xinrui Liu, and Xinyu Li</i>	
Internet of Things Applications in Bulk Shipping Logistics: Problems and Potential Solutions . . . . .	565
<i>Xin Song, Lei Huang, and Stefan Fenz</i>	
Design and Research of Urban Intelligent Transportation System Based on the Internet of Things . . . . .	572
<i>Hong Zhou, Bingwu Liu, and Donghan Wang</i>	
A Single Sign-On Scheme for Cross Domain Web Applications Based on SOA . . . . .	581
<i>Enze He and Qiaoyan Wen</i>	
A Fast FCD-Based Dynamic Traffic Navigation System . . . . .	590
<i>Shusu Shi, Zonghui Wang, Licheng Yu, and Wenzhi Chen</i>	
Design of Field Information Monitoring Platform Based on the Internet of Things . . . . .	597
<i>Keqiang Wang and Ken Cai</i>	
Entropy-DEA Evaluation of Agile Supply Chain Management Based on IOT . . . . .	603
<i>Xiaowen Wang, Dihui Lai, Jinsheng He, and Shu'en Wang</i>	
AST-Based Plagiarism Detection Method . . . . .	611
<i>Liping Zhang, Dongsheng Liu, Yanchen Li, and Mei Zhong</i>	
A Research on Plagiarism Detecting Method Based on XML Similarity and Clustering . . . . .	619
<i>Shengying Jia, Dongsheng Liu, Liping Zhang, and Chenglong Liu</i>	
Research on the Relevant Standards of Internet of Tings . . . . .	627
<i>Jie Jing and Hao Li</i>	

XVI Table of Contents

The Computational Intelligence of Computer Go ..... <i>Bin Wu</i>	633
Visualizing the Random Forest by 3D Techniques ..... <i>Min Yang, Hexin Xu, Dingju Zhu, and Huijuan Chen</i>	639
The Computational Intelligence of the Game Pac-Man ..... <i>Bin Wu</i>	646
A Reliable AOP Technique Using XML to Handle Semantic Aspect Problems ..... <i>Eunsun Kim and Byungjeong Lee</i>	652
Design of a Smart Meter Recorder with Mass Storage Based on DL/T645-2007 Protocol ..... <i>Jia Guo and Dong Liu</i>	660
Research on Temperament Classification Method Based on Fuzzy C-Means ..... <i>Wenjun Yang, Yingying Zhu, and Yongxin Li</i>	667
<b>Author Index .....</b>	<b>673</b>

# Wireless Sensing and Propagation Characterization for Smart Greenhouses

Yuyang Peng, Ushan He, and Jaeho Choi

Dept. of Electronic Engineering, CAIIT, Chonbuk National University

Chonju, Chonbuk, Rep. of Korea, 561-756

{yuyang, wave}@jbnu.ac.kr, colinmengyu@hotmail.com

**Abstract.** Recently, wireless sensor networks draw much attention as the sensor nodes equipped with various monitoring capabilities have additive advantages over the traditional communication technologies. In this paper, we propose a wireless sensing method for monitoring temperature and humidity of a smart greenhouse in a real time process. The propagation characteristics of communications in a greenhouse have been investigated first in terms of transmission and reception signal strength and signal reception rate. Those properties are further used to characterize the propagation property that can properly represent the communication environment of a greenhouse. The proposed propagation characterization not only well represents the pass loss in the greenhouse but also provides valuable insights and basic guidelines when designing a sensor network for a smart greenhouse.

**Keywords:** Wireless sensor network, Smart greenhouse, Propagation characterization.

## 1 Introduction

The wireless sensor networks (WSNs) is drawing much attention nowadays and expecting to be one of the key technologies in the future. WSNs can operate in a wide range of environments and provide advantages in cost, size, power, flexibility and distributed intelligence [1]. Moreover, WSNs can be deployed anywhere without much intervention with the incumbent communication infrastructures. These networks consist of sensor nodes distributed over a certain area; and the sensor nodes co-operate each other to perform various tasks such as detecting and monitoring. The wireless sensor modules are basic building blocks and they are consisted of micro-controller, radio transmitter, and receiver along with a particular sensing capability [2][3].

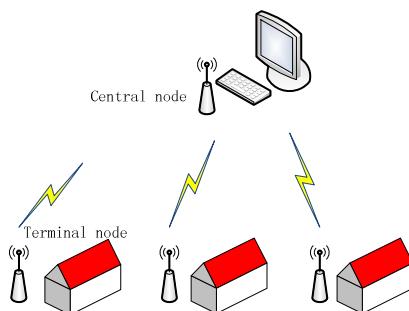
The advantages for deploying a wireless sensor network become apparent in the case of greenhouses where cabling work is unsuitable and costly. Wind power plantation and warehouses are other examples where the choice of WSN deployment is highly feasible. In particular, smart greenhouses need timely ecological and atmospheric information; and the wireless sensors can be placed around the greenhouse to autonomously work themselves to detect and monitor the temperature and humidity of the greenhouse. Above all, the smart greenhouse is much depended

on robust connectivity connecting sensors to deliver the data to the sink of the WSN. In other words, in order to make sure of reliable data transmission over the wireless sensors, the signal transmission characteristics of the sensors should be well defined. Currently, there is one suitable standard technology that we can use to implement a WSN for the greenhouse. It is IEEE802.15.4/ZigBee. It is a low-powered, short distance wireless protocol defined for the physical layer and the medium access control [4][5]. Realizing this standard wireless protocol, there are also commercial products available in the market. Even then, the transmission properties of such sensor modules around the greenhouse are not readily available and they need to be investigated and defined.

The paper is organized as follows: The section 2 discusses the WSN organization and the basic hardware structure of the sensor modules. Section 3 presents the description on greenhouse communication environments and the procedures for collecting greenhouse data. The experiments are performed in Section 4 and the transmission properties are analyzed and formulated in Section 5. Finally, the conclusion is made.

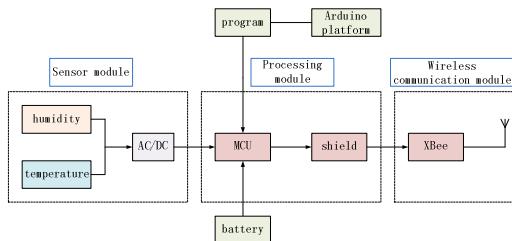
## 2 Sensor Network and Sensor Node

Nowadays smart greenhouse is more than a buzzword. It implies the total integration of automated monitoring technology to carry out time to time operation. The components of a smart greenhouse may include temperature and humidity monitoring system especially in flower and vegetable greenhouse. A smart greenhouse will help a greenhouse manager collect and monitor the environment information and then reduce operating costs. The architecture of smart greenhouse system is shown in Fig. 1. This system is composed of terminal sensor nodes, central node and personal computer. Each terminal node has unique address which can be set by the user. The main task of terminal nodes is to collect data in smart greenhouse environment, and transmit data to the central node. The central node takes the responsibility of receiving information, communicating with the server.



**Fig. 1.** Architecture for smart greenhouses

Now, let us consider the architecture of sensor module is shown in Fig. 2. It includes three modules, i.e., processing module, wireless communication module, and sensing module. First, considering sensor resource requirement and communication reliability, the Atmega-328 processor is adopted in the processing module. Second, for the communications module, the XBee unit produced by Digi is the choice here. It is a ZigBee/IEEE 802.15.4 compliant solution for WSN and it operates in the 2.4 GHz band with the effective data rate of 250 kbps. The XBee is shown in Fig. 3. The main feature and specifications of the module are described in the next table [6]. Third, for the sensing module, a SHT11 unit produced by Sensiron is used. It is a single-chip with a multi-sensing module that can deliver humidity and temperature as calibrated digital outputs [7]. The SHT11 unit is shown along with XBee Fig. 3.



**Fig. 2.** Modules for a sensor node



**Fig. 3.** XBee module and SHT sensing module

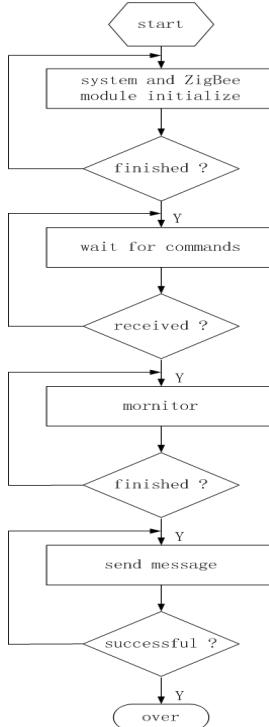
**Table 1.** Parameters of XBee module

Xbee module parameters	
specification	XBee
Transmit power output	1 mW/0dBm
RF data rate	250 kbps
Power requirements	
Supply voltage	2.8–3.4 V
Transmit current	45 mA
Receive current	50 mA
Power down current	<10 µA

### 3 Procedures for Wireless Sensing

The SHT-11 sensor output is connected to the analog to digital converter. The temperature and humidity values are sensed and converted to a 14-bit digital format

which is then fed to the input port of the Atmega-328. The Arduino programming language can be used to prepare source codes. The codes mainly include procedures for detecting temperature and humidity and then transmitting them to the monitoring center. The flow chart for these procedures is shown in Fig. 4.



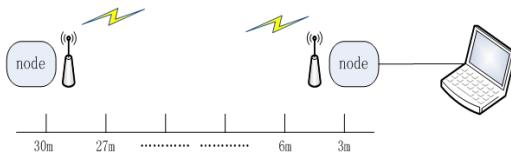
**Fig. 4.** Flow chart for the sensor node operation

- Step 1: Initialize the module by turning on the power of the sensor node.
- Step 2: After initialization, the sensor waits for a command message.
- Step 3: If the sensor received a command, it monitor temperature and humidity. Otherwise, it will go back to Step 2.
- Step 4: At the end of monitoring, it transmits data to the monitoring center.
- Step 5: If the transmission is unsuccessful, it sends them again. Otherwise, it goes back to Step 3 for the next command.

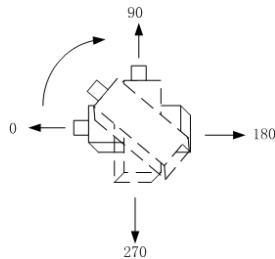
## 4 Field Experiments

Experiments have been conducted in order to verify the performance and reliability [8] of the proposed sensor node module. They are carried out in an experiment playground. X-CTU software and two modules are used in this experiment. One acts as the based station which records the received signal strength of all the packets at a fixed transmitting output power. The distance between the source sensor node and the base

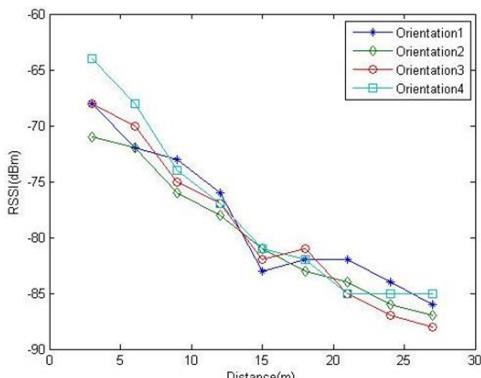
station varies as shown in Fig. 5. For each distance measuring the received signal strength four different source node orientations are used as illustrated in Fig. 6. The source sensor node is placed 0.32 meters above the ground with 0 dBm transmitting power. As shown in Fig. 7, four curves representing four different source node orientations, respectively, show a similar downward trend as the distance between two entities increases as expected. On the other hand, Fig. 8 shows the relationship between the packet delivery rate and the received signal strength. At the received signal strength of -82 dBm or above, the packet reception rate maintains at least 80 percent. However, the packet delivery rate becomes unstable when the received signal strength weakens below -82 dBm. Finally, Fig. 9 shows that how the link quality varies over the distance between two sensor nodes operating in an open space [9]. The experiment results show that there is a range of distance over which two nodes can sustain a good connectivity. The transmission power and the transmission distance affect the successful reception rate; and in case of the smart greenhouse this measure should be carefully approached in order to maintain an errorless humidity and temperature control.



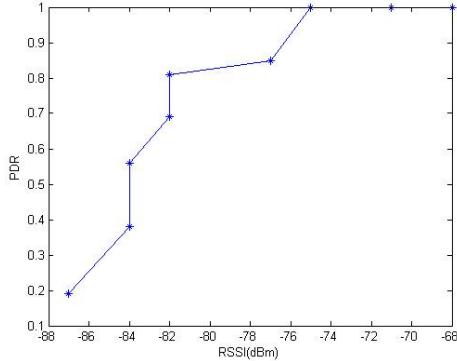
**Fig. 5.** Varying distance between the source sensor node and the base station node



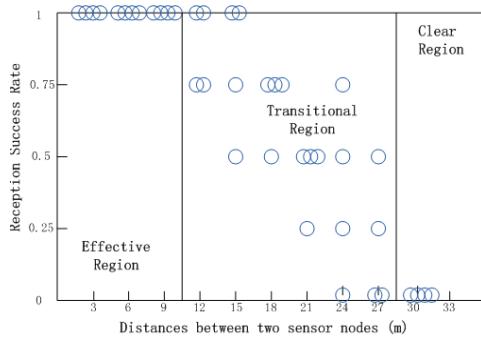
**Fig. 6.** The source sensor node orientations



**Fig. 7.** Received power for each orientation of the source node with respect to the distance



**Fig. 8.** The packet delivery ratio with respect to the received power



**Fig. 9.** The successful reception rate with respect to the distance

## 5 Greenhouse Propagation Characterization

The signal propagation characterization for the smart greenhouse can be divided into two parts, i.e., the open-space part and the foliage part, and it can define as follows:

$$PL = PL_o + PL_f \quad (1)$$

where  $PL_o$  is the open-space part and  $PL_f$  is the foliage propagation loss part.  $PL_o$  is calibrated with the data collected from the experiments performed and it follows a log-distance model as below:

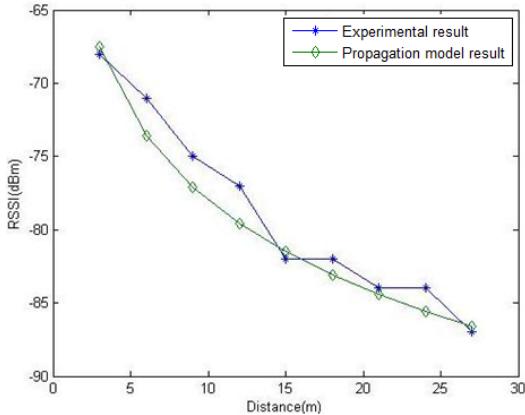
$$PL_o = p_i + n \log_{10}(d) \quad (2)$$

where  $p_i$  is the initial path loss within 1m radius,  $d \geq 1m$  is the distance from the source sensor node to the base station node, and  $n$  is the path loss exponent with the value of 2 in a normal case. In the smart greenhouse, additively, the attenuation rises due to the foliage. From the existing indoor propagation characterization study, the foliage induced excess loss is as follows [10]:

$$PL_f = 0.2f^{0.3}d^{0.6} \quad (3)$$

where  $f$  is the frequency in GHz, and  $d$  is the distance from the source node. Hence, Eq. (1) can be rewritten as follows:

$$PL = p_i + n \log_{10}(d) + 0.2f^{0.3}d^{0.6} \quad (4)$$



**Fig. 10.** The propagation characterization curve along with the empirical curve

Fig. 10 shows the two curves. One is from experimental data and the other is the numerical one plotting Eq. (4). The experimental data fall very near to the proposed characterization curve.

## 6 Conclusions

The ZigBee technology based sensor module for temperature and humidity monitoring for the smart greenhouse has been designed, implemented, and verified. Also, the propagation characterization is presented for the smart greenhouse. The experiment results shows that the proposed propagation characterization curve can provide an average pass loss in the greenhouse. The proposed model here is simple and realistic; thus it is useful to save the cost for building a wireless sensor network suitable for a smart greenhouse. The future research is continued to investigate the behavior of the sensor network when the topology changes due to blocking.

**Acknowledgement.** This work has been supported by the University Research fund provided by Chonbuk National University 2010.

## References

1. Ruiz-Garcia, L., Barreiro, P., Robla, J.: Performance of ZigBee-Based wireless sensor nodes for real-time monitoring of fruit logistics. *Journal of Food Engineering* 87(8), 405–415 (2008)
2. Frran, P., Flammini, A., Marioli, D.: IEEE802.11 Sensor Networking. *IEEE Trans. Instr. and Means.* 55(2), 615–619 (2006)

3. Qiu, P., Heo, U., Choi, J.: A gateway protocol architecture for Zigbee based wireless sensor network interconnecting TCP/IP networks. *Jr. of KISPS* 10(3), 176–180 (2009)
4. ZigBee Specification 1.00, ZigBee Alliance (2004)
5. Ergen, S.: ZigBee/IEEE 802.15.4 Summary. University of California Berkeley (2004)
6. XBee/XBee-Pro OEM RF Modules, Product Manual v 1.0, MaxStream (2005)
7. XBee/XBee-Pro OEM RF Modules, Product Manual v 1.0, MaxStream (2006)
8. Tang, L., Wang, K., Huang, Y., Gu, K.: Channel characterization and link quality assessment of IEEE 802.15.4-compliant radio for factory environments. *IEEE Trans. Industrial Info.* 3(2), 99–110 (2007)
9. Lu, B., Gungor, V.: Online and remote motor energy monitoring and fault diagnostics using wireless sensor networks. *IEEE Trans. Ind. Electron.* 56(11), 4651–4659 (2009)
10. Meng, Y., Lee, Y., Ng, B.: Empirical Near Ground Path Loss Modeling in a Forest at VHF and UHF Bands. *IEEE Trans. Antennas Prop.* 57(5), 1461–1468 (2009)

# Efficient Ad Hoc Routing Technique for Connecting Geographically Distributed Heterogeneous Networks

Youngshin Ahn, Yi Zhao, Ushan He, and Jaeho Choi

Dept. of Electronic Engineering, CAIT, Chonbuk National University  
Chonju, Chonbuk, Rep. of Korea, 561-756

jwshan@hanmail.net, colinmengyu@hotmail.com, wave@jbnu.ac.kr

**Abstract.** Connecting a mobile ad hoc network and wireless sensor network is the focus of our investigation. In order to remotely access a sensor node from a mobile node, first, a node in the MANET needs to be connected to a sensor network coordinator. The challenge is to consider node mobility and limited sensor battery power. Here, an indication metric determined based on the MANET local information is designed to provide the link lifetimes and helps a mobile node to adaptively update its link and the route. Furthermore, the paper describes an energy-saving tree construction scheme for connecting sensors in the sensor network. The performance of the proposed method is verified using OPNET. The results in terms of response time and packet delivery ratio with respect to the number of sensor network coordinators show that the proposed connection method is superior to the conventional method.

**Keywords:** Ad hoc routing, MANET, Sensor networks, Heterogeneous networks, Data delivery.

## 1 Introduction

Current technologies require a fixed infrastructure for communication between mobile terminals. However, in some cases, such as military operations or rescues, a fixed infrastructure is not always available.

Moreover, it is sometimes desirable to program mobile nodes to solve a variety of problems within specific geographical areas without relying on an infrastructure. In some applications, a mobile node may need to collect data from a specific region. These factors leads to draw interests in studying sensing related problems in infrastructureless networks [1][2].

The ad hoc remote sensing has its origin in the conventional connecting method in which packets can be sent to some member nodes within a specified geographical region [3]. In a traditional multicast, if a node wants to be a member of a certain multicast group, it has to explicitly join the group. However, in the traditional method, a node automatically becomes a member if it is in the geographical region [4].

There are several such ad hoc routing protocols proposed recent years. Those include the location based multicast algorithm (LBM) [5], GeoGRID [6], and GeoTORA [7]. More literatures related to the ad hoc multicasting can be found in [8-12]. Mostly, these

protocols consist of two stages. The first stage is routing from the source to the geographical region; and the second stage is data aggregation within the geographical region. They often concentrate on routing issues outside of the geographical region; and their protocols attempt to identify the best route from the source node to the region.

The investigation on the ad hoc routing technique proposed in this paper involves two heterogeneous networks. The proposed method aims to establish an efficient ad hoc route over the mobile ad hoc network (MANET) using the indication metric computed from the local information; and also aims to construct an energy-efficient sensor tree in the sensor region by computing the battery power balance equation.

The remainder of this paper is organized as following. Section 2 presents the proposed ad hoc routing technique over the MANET; and also describes the sensor tree construction scheme in the sensor region. In Section 3 the simulation scenarios and simulation results are discussed. Finally, the conclusions are made in Section 4.

## 2 The Proposed Ad Hoc Routing Technique

The mobile ad hoc region and the sensor region are two heterogeneous network regions involved in the proposed ad hoc routing method. A route between two heterogeneous regions to be established and its process consists of two phases.

### 2.1 Routing in the MANET

#### A. Hello Message

A new hello message protocol (HMP) is proposed for the MANET in order to cope with node mobility. Most MANET routing protocols are based on hello beaconing, and it is important and meaningful to improve its efficiency.

The HMP presented here is designed in such a way to dynamically adapt to the relatively speeds between the source node and its neighboring nodes. If two nodes move with a relative speed  $S$ , the broadcasting frequency of the HM can be defined as follows [13]:

$$F_{opt} = \frac{2S}{\alpha R} \quad (1)$$

where  $\alpha R$  is the transmission distance and is  $\alpha$  a constant,  $0 \leq \alpha \leq 1$ .

However, if there are more than two nodes within the coverage area of the source node, a new equation for the HM beaconing frequency can be defined as follows:

$$\bar{F} = \frac{2 \sum_{j=1}^n S_{ij}}{\alpha n R} (i) \quad (2)$$

where  $S_{ij}$  is the relative speed between the observing Node  $i$  and its neighboring Nodes  $j$  assuming there are  $n$  neighboring nodes. If the relative speeds tend to slow

down, the HM beaconing frequency also decreases and, hence, saves the energy consumption of the nodes.

### B. Link Lifetime

Consider two neighboring nodes in the MANET. The time duration for which these two nodes are connected each other is called the link lifetime.

Assume that  $(x_i, y_i)$  is the coordinate of the mobile Node  $i$  and  $(x_j, y_j)$  is the coordinate of the mobile Node  $j$ . In addition, let  $v_i$  and  $v_j$  be the velocities and  $\theta_i$  and  $\theta_j$  be the moving directions of Nodes  $i$  and  $j$ , respectively. Then, the amount of time that they will be connected each other is defined as follows:

$$LT(i) = \frac{-(ab + cd) + \sqrt{(a^2 + c^2)r^2 - (ad - bc)^2}}{a^2 + b^2} \quad (3)$$

where  $r$  is the transmission range and

$$\begin{aligned} a &= v_i \cos \theta_i - v_j \cos \theta_j \\ b &= x_i - x_j \\ c &= v_i \sin \theta_i - v_j \sin \theta_j \\ d &= y_i - y_j \end{aligned}$$

### C. Route Discovery

In the proposed MANET routing technique, each mobile node periodically exchanges HMs in which its coordinates, own ID, and neighboring node IDs are appended. Upon receiving a HM, the each node calculates the link lifetime between itself and the neighboring nodes using Eq. (3). In addition to the link lifetime, the nodes periodically calculate the good link indication metric as follows:

$$N(i) = \alpha LT(i) + \beta EL(i) + \gamma / D(i) \quad (4)$$

where  $\alpha, \beta$  and  $\gamma$  represent the weights and  $0 < \alpha + \beta + \gamma \leq 1$ ;  $LT(i)$  is the link lifetime;  $EL(i)$  is the energy level; and  $D(i)$  is the distance. Upon receiving route request from a source node, the intermediate nodes located toward the destination can use the vital link with the highest  $N(i)$  value. Depending on the applications, the values for the weights can be adjusted. For the services requiring real-time, for example, one can set the weight  $\gamma$  relatively heavy. Then, it can quickly identify the target route and transfer the packets as soon as possible with a minimum delay assuming all other conditions are same.

When a link failure occurs, a new link needs to be discovered. The source node needs to initiate a route rediscovery procedure if a timeout occurs on the current route. In contrary to the conventional ad-hoc routing, the proposed MANET routing technique based on the novel HMP and  $N(i)$  can improve the route recovery upon a route failure.

## 2.2 Sensor Tree Construction in the Sensor Region

In the proposed ad hoc routing algorithm, the second phase proceeds with constructing a sensor tree joining together sensors distributed over a certain geographic area. Here, any number of coordinators deployed in the sensor region can play a role as a sink as well as a gateway.

In order to collect data from the sensors, a tree is constructed for each coordinator. At first, the sensors need to exchange initialization hello messages (IHM) by broadcasting IHM to their one-hop neighbors. The IHM includes the geographic coordinates and unique ID of the sending node. After exchanging the first IHM, the coordinator knows the identities and locations of its one-hop neighbors. Then, every node broadcasts the second IHM to its one-hop neighbors, including its ID, coordinates, and the number of one-hop neighbors. After exchanging the second IHM, every node knows the identities and locations of its neighboring nodes up to two hops.

Based on the information obtained through IHM exchanges, the tree construction begins from the coordinator. It calculates the powers of the neighbor nodes and chooses a node with the lowest power increment as the forwarding node. The chosen forwarding node will then join the sensor tree where it will modify the forwarding power and repeat the forwarding node choosing steps to construct a sensor tree until all possible sensor nodes join the tree.

For the above mentioned incremental power calculation involved in choosing a forwarding node, the average broadcasting power is used as a power measurement and it is defined as follows:

$$ABE(i) = P(i) / M(i) \quad (5)$$

where  $P(i)$  is the power available in Node  $i$  to reach the farthest one-hop neighbor node, and  $M(i)$  is the number of one-hop neighboring nodes which are not yet included in the tree. In choosing a next-hop tree node, the tree-edge Node  $i$  calculates  $ABE(i)$  and chooses the free node with the smallest  $ABE(i)$  value as its next-hop node.

Once a sensor tree for a coordinator is completed, another coordinator can repeat the tree construction steps until every coordinator builds its own sensor tree. In our method, three coordinators are used considering complexity of addressing and overhead. Upon completion of sensor tree construction, the sensor nodes undergo their normal operation mode and periodically they transmit beacons to their one-hop neighbors to inform their existence and to determine the identities of their failed neighbor nodes, if any.

When a request packet issued from the source node in the MANET arrives at the coordinator of the sensor network, depending on the request, the AP can respond with its stored data or wake up sensors in the sensor tree to collect and provide the updated sensor data as a reply to the request.

## 3 Computer Simulations

A set of computer simulations is performed using the OPNET modeler. The actual sizes of the network areas can be varying, however, in the simulation scenarios the scaled-down network areas are considered. The MANET region covers most area

while the sensor network region covers the upper right corner. In the MANET, 20 mobile nodes move around with a speed of 5 ~ 20unit/sec and their mobility is defined by the random way-point model. IEEE 802.11 DCF is used to control the medium access among the mobile nodes. On the other hand, 50 Zigbee sensors are planted uniformly in the upper right corner.

The performance of the proposed routing technique is evaluated in terms of packet delivery ratio and end-to-end response time. Specifically, the packet delivery ratio is defined as the ratio between the number of correctly received data packets and the total number of packets sent; and the end-to-end response time is defined as the time duration from a departure of a request by a source node to an arrival of a reply from a coordinator.

For the first set of performance evaluation, the packet delivery ratio is observed and compared with the LBM method. As shown in Fig. 1, for the proposed method, the packet delivery ratio is up to 95% after 100 seconds of simulation time. At the initial part of the simulation time, however, the proposed method under-performs due to route and tree establishment. Going into a normal operation, though, the packet delivery ratio curve for the proposed method settles above that of the LBM. Note that for the LBM, there is no fixed forwarding tree for the sensors in the sensors region.

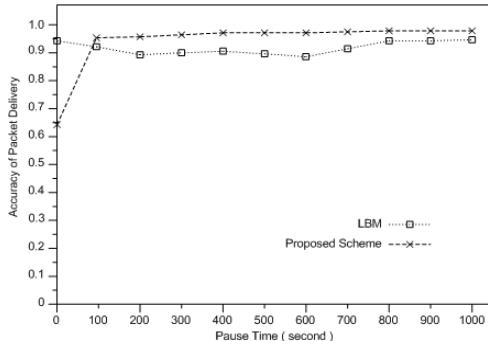
For the second set of simulations, the end-to-end response time is observed and compared to that of the LBM. The curves for the proposed method and the LBM are shown in Fig. 2. The time difference comes within 0.003 seconds, and the curve for the proposed method lies mostly below the LBM curve representing a bit shorter response time. Similarly to the curves in Fig. 1, the response time is longer at the beginning due to tree building initiation processes.

For the third set of simulations, the behavior of the proposed scheme in term of the mobility of mobile nodes is observed. As shown in Fig. 3, two curves represent the packet delivery ratios for the proposed method and the LBM, respectively. As expected, the packet delivery ratios drop for both as the mobility increases. However, the performance degradation is sharper for the LBM than that of the proposed method. Even at the higher speed, however, the packet delivery performance for the proposed method degrades gracefully down around to 90 percent. This result supports the usefulness of the proposed HMP, which is designed to work adaptively to the mobile node mobility and to provide a good link with a relatively longer link lifetime.

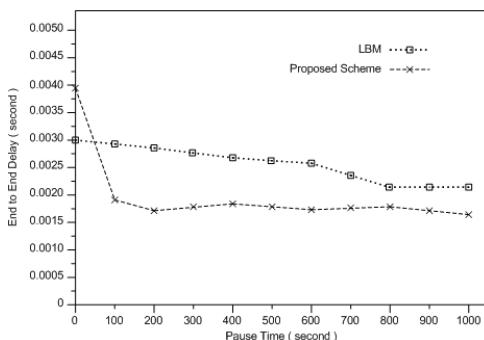
For the fourth set of simulations, the battery power consumption is observed and also compared to that of the LBM. Fig. 4 shows the average remaining battery power level in each node. In the initial phase, since it requires to broadcast HMP packets to set up a route in the MANET and also IHM packets to set up a tree in the sensor region, respectively, the proposed method consumes more battery power; and therefore, the curve drops sharply at the beginning. However, once coming into the stabilized operation mode, the proposed method is more effective in saving battery power than that of the LBM. After 10 minutes of simulation time, around 60 percent of battery power is consumed for the proposed method while 70 percent is consumed for the LBM.

For the final set of simulations, the response time is measured with respect to the number of coordinator APs in the sensor network region. Fig. 5 shows the curves for three different cases: the top curve is for the one-AP case; the middle one for the two-coordinator APs case; and, the bottom one the for 3-coordinator APs case. As the

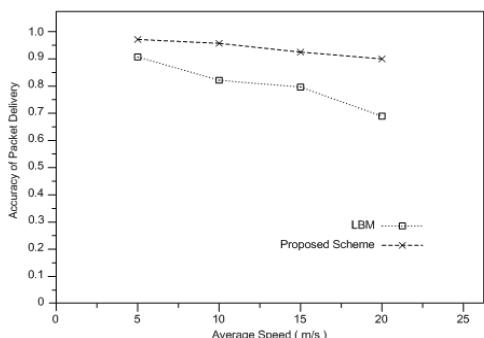
results shows the response time gets quicker with respect to the rising number of coordinator APs; and the response time gets quicker as much as two-times quicker by using three coordinator APs instead of just one. It should be noted, however, that multiple coordinator APs can increase the complexity and cost in deployment and implementation.



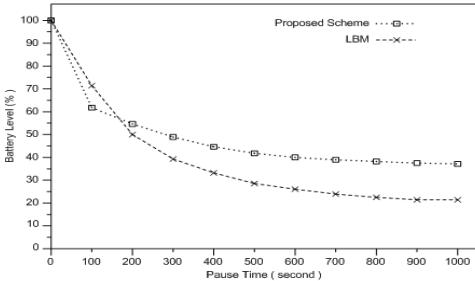
**Fig. 1.** Performance in packet delivery ratio



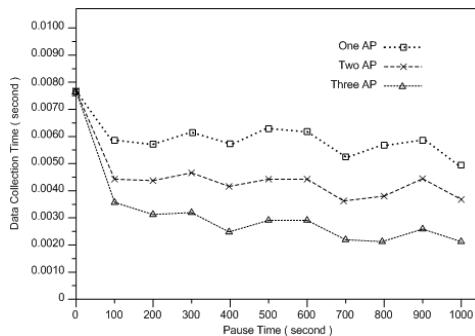
**Fig. 2.** Performance in end-to-end response time



**Fig. 3.** Mobility effect on packet delivery ratio



**Fig. 4.** Comparison on battery power consumption



**Fig. 5.** Coordinator APs effect on response time

## 4 Conclusions

In this paper we have presented a novel ad hoc routing technique for connecting two heterogeneous networks, i.e., MANET and WSN. Since the wireless links are volatile in nature, a new HMP is proposed so that a mobile node can effectively discover a link on a fly. Furthermore, we have incorporated an energy efficient sensor tree construction algorithm for the sensors in the sensor region. The energy balancing parameter called the average packet broadcasting power proved to elongate the lifetime of the sensors network. The simulations results show us that the proposed method is better than the conventional method such as LBM in terms of packet delivery ratio and end-to-end response time. Also, multiple coordinators deployed in the sensor region can be much effective to improve the performance at the cost of increased addressing and implementation complexity.

**Acknowledgement.** This work has been supported by the University Scholarship fund provided by Chonbuk National University 2012.

## References

1. Kim, S., Jung, Y., Lee, S., Lee, E., Ahn, S.: Efficient Geocast Utilizing Topology Information Database. In: Proc. of CIT Workshop, pp. 210–215 (2008)
2. Zhao, Y., Heo, U., Qiu, P., Wan, X., Choi, J.: An Efficient Tree-Based Scheme for Geocast Protocol in Heterogeneous Wireless Sensor Network. In: Proc. of ICITA 2009, pp. 228–233 (2009)
3. Yao, P., Krohne, E., Camp, T.: Performance Comparison of Geocast Routing Protocols for a MANET. In: Proc. of IEEE ICCCN, pp. 213–220 (2004)
4. Lee, S., Ko, Y.: Geometry-Driven Scheme for Geocast Routing in Mobile Ad Hoc Networks. In: Proc. of Veh. Tech., pp. 638–642 (2006)
5. Ko, Y., Vaidya, N.: Geocast in Mobile Ad Hoc Networks: Location-based Multicast Algorithms. In: Proc. of IEEE WMCSA, New Orleans, pp. 101–110 (1990)
6. Liao, W., Tseng, Y., Lo, K., Sheu, J.: GeoGRID: A Geocast Protocol for Mobile Ad Hoc Networks based on GRID. Jr. of Internet Tech. 1(1), 23–32 (2000)
7. Ko, Y., Vaidya, N.: GeoTORA: A Protocol for Geocast in Mobile Ad Hoc Networks. In: Proc. of ICNP, pp. 213–220 (2004)
8. Stojmenovic, I.: Voronoi Diagram and Convex Hull Based Geocast and Routing in Wireless Networks: Tech. Report, TR-00-11, University of Ottawa (1999)
9. Ko, Y., Vaidya, N.: Flooding-Based Geocast Protocols for Mobile Ad Hoc Networks. Mobile Networks and Applications Journal 7(6), 471–480 (2002)
10. Eason, G., Noble, B., Sneddon, I.: Secure and Energy Efficient Geocast Protocol for Sensor Networks. In: Proc. of Info. Tech. and Applications, pp. 214–219 (2008)
11. Kummakasikit, M., Thipchaksura, S., Varakulsiripunth, R.: Performance Improvement of Associativity-Based Routing Protocol for Mobile Ad Hoc Networks. In: Proc. of Inform. Comm. and Signal Processing, pp. 16–20 (2005)
12. Wieselthier, J., Nguyen, G., Ephremides, A.: On the Construction of Energy-Efficient Broadcast and Multicast Trees in Wireless Network. In: Proc. of IEEE Infocom 2000, pp. 585–594 (2000)
13. Troël, A.: Prise en compte de la mobilité dans les interactions demobilité entre terminaux à profils hétérogènes, doctoral dissertation, Université de Rennes 1, France (2004)

# An Adaptive Scheduling Algorithm for the Patient Monitoring System on WBANs

Hongkyu Jeong

Department of High Tech. Medical System, Kyungil University  
hongkyu.jeong@gmail.com

**Abstract.** The Patient Monitoring System (PMS) is an essential medical system to check the vital signs of patients for 24 hours. However, the wire-line connection makes patients uncomfortable and makes it inconvenient for medical doctors and nurses to manage the system. Recently, Wireless Body Area Networks (WBANs) has been spotlighted as a promising wireless communication technology for the devices operating on, in or around the human body. Thus, WBAN could be a candidate technology to resolve the connection problem of the PMS. However, existing WBAN technologies, especially the IEEE 802.15.4 and the IEEE 802.15.6, have some limitations to support the PMS in the viewpoint of time sensitivity and energy efficiency. Therefore, this paper proposes an Adaptive Scheduling Algorithm (ASA), which supports a fast retransmission mechanism through adaptive retransmission slot allocation within the superframe. In addition, the ASA piggybacks the synchronization and association control data into existing ACK in order to save wireless resources. Through the performance comparison, the proposing ASA reduces the average latency, enhances energy efficiency, and minimizes the average link occupancy compared to the IEEE 802.15.4 and the IEEE 802.15.6 technologies.

**Keywords:** wireless body area network, patient monitoring system, adaptive scheduling algorithm.

## 1 Introduction

The rapid growth in physiological sensors, low power integrated circuits and energy-efficient wireless communication has enabled a new generation of Wireless Body Area Networks (WBANs). As a WBAN technology, the IEEE 802.15.4 is utilized by controlling power and adapting network structure to star topology [1]. However, it supports a low data transmission rate and low-level Quality of Service (QoS) mechanism. Most importantly, it has a limitation from the viewpoint of the energy-saving mechanism to satisfy the requirements of WBAN. In order to overcome the limitations of the IEEE 802.15.4, the IEEE 802.15.6 technology is proposed with a sufficient QoS mechanism and high transmission rate up to 10 Mbps [2].

The Patient Monitoring System (PMS) can be a killer application applicable for WBAN since patients are often uncomfortable with the wire-lines of the

PMS and medical doctors and nurses (i.e., medical team) have some troubles to manage the PMS because of the wire-lines. The PMS displays the vital signs of various applications simultaneously such as ECG, EEG, EMG, O<sub>2</sub>/CO<sub>2</sub> rate, temperature, and blood glucose. If a vital sign is detected at an abnormal state based on the predefined threshold value, an alarm will sound. Therefore, the vital sign data of the PMS should be transmitted with a strict time constraint. Moreover, the energy consumption of slave nodes should be minimized since the medical team cannot afford to change the battery of slave nodes every day. From these viewpoints, the IEEE 802.15.6 has some drawbacks to support the PMS. This is the motivation of this paper. We propose an Adaptive Scheduling Algorithm (ASA), which supports a fast retransmission mechanism via adaptive retransmission slot allocation within the superframe. In addition, the ASA does not require a beacon frame for synchronization and association control but utilize existing ACK by piggybacking such information.

The following sections of this paper are organized as follows. In Section 2, we analyze the related technologies, the IEEE 802.15.4 MAC and IEEE 802.15.6 MAC. In Section 3, we present our proposal with an operational mechanism. In Section 4, we evaluate the performances of the proposed ASA compared to the existing candidate MAC protocols. Then we conclude the paper in Section 5 with our contribution.

## 2 Related Works

In this section, we describe the IEEE 802.15.4 and the IEEE 802.15.6 standard technologies briefly, which are commonly utilized for Wireless Body Area Networks (WBANs)[\[1,2,3,4,5,6\]](#).

### 2.1 IEEE 802.15.4

The IEEE 802.15.4 describes the physical layer and MAC layer for low data rate, short-range Wireless Personal Area Networks (WPANs). However, the IEEE 802.15.4 standard technology is commonly used for WBANs by controlling power and adapting network structure to star topology [\[1,2,3,4\]](#).

In the beacon enabled mode of the IEEE 802.15.4, slave nodes competitively try to send a frame to the coordinator node in order to reserve Guaranteed Time Slots (GTSs) in Contention Free Period (CFP) and each slave node is awakened during active periods. On the other hand, slave nodes turn off their radio modules to minimize power consumption in the inactive period. If the frame transmitted by a slave node does not arrive or arrives with some errors at the coordinator node, the frame should be retransmitted. For this, the coordinator node allocates more slots of the CFP to the slave node in the next superframe. On the contrary, in the non-beacon enabled mode the IEEE 802.15.4 makes use of Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) to transmit frames in order to avoid frame collision like the CAP of beacon-enabled mode.

The IEEE 802.15.4 MAC has several limitations to allocate wireless resources multiple times for a slave node in a superframe, to save energy in slave nodes

because of relatively long wake-up duration, and high link occupancy because of the mandatory CAP period.

## 2.2 IEEE 802.15.6

Although the IEEE 802.15.4 standard has been spotlighted as a candidate technology for implementing WBAN for a decade, it contains several limitations. In order to overcome these limitations, the IEEE 802.15.6 standard is proposed.

The IEEE 802.15.6 MAC [5,6] has several characteristics: firstly, it supports three different access modes, which are beacon mode with a superframe, non-beacon mode with a superframe, and non-beacon mode without a superframe. Secondly, it provides various ACK methods such as Immediate ACK (I-ACK), Group ACK (G-ACK), Block ACK (B-ACK), Late ACK (L-ACK), and No ACK (N-ACK). Thirdly, in a beacon mode it supports the control of sleep and wake-up interval based on the superframe. Lastly, it provides user priority according to the traffic type and/or frame type, so that high priority traffic such as medical data or emergency data has many opportunities to be delivered.

Most of all, in the IEEE 802.15.6 MAC supporting the beacon mode access mode a coordinator node sends a beacon frame per every superframe. Each beacon frame provides information of energy management for each node, synchronization, and resource allocation. Access periods include Exclusive Access Phase (EAP) for emergency event data, Random Access Phase (RAP) for emergency events, Managed Access Phase (MAP) for scheduled uplink/downlink allocation intervals, and Contention Access Phase (CAP) for contended transmission with a random back-off period.

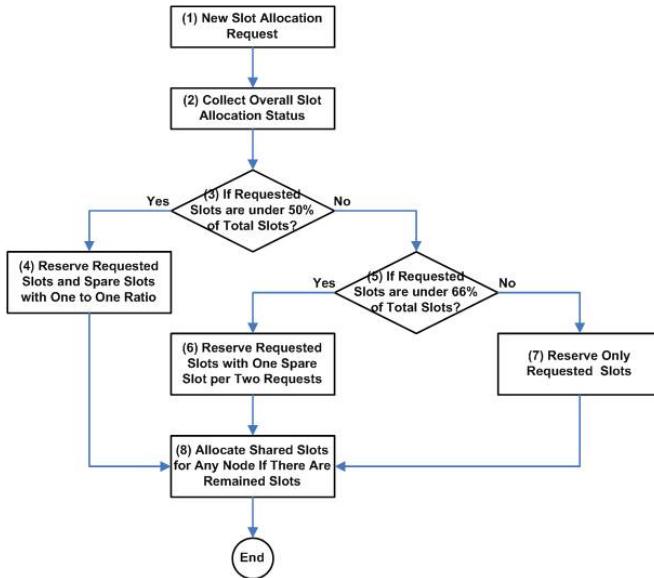
Although the IEEE 802.15.6 MAC improves energy-saving operation compared to the IEEE 802.15.4 MAC, it still has weaknesses in the viewpoint of the average latency because of distinguished access periods (e.g., EAP and MAP) for each purpose. Moreover, the IEEE 802.15.6 MAC cannot support improvised resource allocation in the case of a data transmission fail if next slots after reserved slots are already reserved for another node.

## 3 Proposing Adaptive Scheduling Algorithm (ASA)

We describe the proposing Adaptive Scheduling Algorithm (ASA) in this section, which achieves lower latency performance, especially in the case of the frame retransmission, greater energy conservation and optimizes the link occupancy rate compared to the aforementioned MAC protocols.

### 3.1 Description of the ASA

As we explained about the PMS at the Section 1, the vital signs of the PMS are time-critical, so the frames containing the vital signs should arrive at the coordinator node with little delay. In addition, the energy of sensor nodes (i.e., slave nodes) should be conserved, so that the medical team does not need to

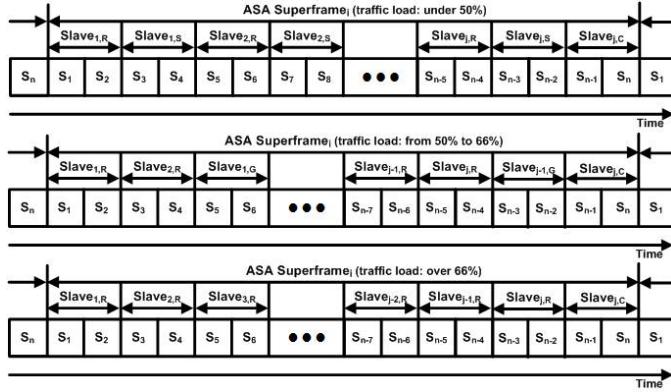


**Fig. 1.** Flow Chart of the Proposing ASA

spend excess time tending to the battery of slave nodes. In order to support various applications simultaneously in a WBAN, link occupancy rate should be minimized.

In order to satisfy the requirements of low latency, low energy consumption, and low link occupancy rate, we propose an Adaptive Scheduling Algorithm (ASA). The ASA has several characteristics: firstly, the ASA operates adaptively based on the traffic load. Basically, the ASA tries to provide spare slots for each requested slots just after the reserved slots, so that it is possible to retransmit the lost data immediately and thereby minimize latency of data transmission completion. Secondly, the ASA does not utilize a beacon frame for the purpose of synchronization and association control. Instead, the ASA adds the information to the ACK (when it is required). Therefore, it is possible to save the energy of slave nodes by reducing awakening time as well as minimize wireless resource consumption. Lastly, the ASA provides contention-free scheduling among slave nodes through reservation-based scheduling with the superframe structure.

Specific algorithm steps of the ASA are described in Figure 1. Threshold values such as 50% and 66% are dependent on the network operators, and the proposed values in Figure 1 are some of the candidate examples. (1) When a new slot allocation for a slave node is requested, (2) the coordinator node collects overall slot allocation status. (3) If overall requested slots (including a new request) are under 50% of the total number of slots in the superframe, (4) the coordinator node reserves both requested slots and spare slots at a one-to-one rate. (5) On the other hand, if overall requested slots are over 50%, it checks whether the allocation rate is over 66%. (6) If the allocation rate is under



**Fig. 2.** Examples of ASA Superframe Structure

66%, it reserves requested slots with spare slots in the ratio of one spare slot per two requested slots. If the overall slot allocation rate becomes over 50% because of the new allocation request, the coordinator node should alter the existing spare slot allocation ratio from one-to-one to one-to-two, and let the slave nodes know about this alteration through the next ACK. At this moment, a new slot allocation is applied after notifying this change to the related slave nodes. (7) Moreover, if overall requested slots are over 66%, the coordinator node only allocates requested slots without a spare slots. (8) In all cases, if there are unscheduled slots, the coordinator node reserves it as shared slots for any slave nodes. If a slave node wants to make use of the shared slots, the slave node utilizes the shared slots based on a CSMA/CA method. In this flow chart, the network operators should determine the adaptive capabilities of the ASA by controlling the threshold parameters.

Figure 2 shows an example of the ASA superframe structure. According to the aforementioned description, each slave node receives spare slots based on the slot allocation requests. If the overall allocation rate is under 50%, each slave node receives requesting slots ( $Slave_{i,R}$ ) with spare slots ( $Slave_{i,S}$ ). In addition, if the overall allocation rate is between a 50% and a 66%, one spare slot ( $Slave_{i,G}$ ) is allocated per two requested slots. Moreover, if overall slave allocation rate is over 66%, the coordinator node only reserves requested slots ( $Slave_{i,R}$ ) and reserves shared slots ( $Slave_{i,C}$ ) if there are remaining slots. The number of requesting slots can be differentiated based on application traffic, but the flow chart and example of the ASA superframe structure just show general cases.

## 4 Performance Evaluation

This section evaluates the performance of the proposed ASA regarding the average latency between the slave nodes and the coordinator node, the remaining

energy state of the slave node, and the average link occupancy compared to the IEEE 802.15.4 MAC and the IEEE 802.15.6 MAC.

#### 4.1 Assumptions

The results were obtained under the following assumption: (i) network topology is a star structure, where the slave nodes are connected to a coordinator node. For the clear performance comparison, a slave node is mainly operated; (ii) an ECG application (shown in the Table 1) is utilized for the performance comparison, where data packet is generated every 10 ms; (iii) the beacon interval of the IEEE 802.15.4 and IEEE 802.15.6 is 50 ms; (iv) in the case of the ASA, the coordinator node schedules the superframe resource based on the traffic load; (v) the remaining energy state and link occupancy of slave nodes are tested in the environment of offered load 0.5; (vi) a 10% loss is considered for all the tests; (vii) the event-driven simulation is run until the remaining energy of the slave node is fully exhausted and repeated for 30 times to obtain the average values. The parameters used in the simulation are summarized in Table 1.

#### 4.2 Performance Comparison

As aforementioned, we evaluated the performance of the proposed ASA from the viewpoints of the average latency, the remaining energy state of the slave node, and the average link occupancy related to the IEEE 802.15.4 MAC and the IEEE 802.15.6 MAC.

Figure 3 shows the average latency of slave nodes. The IEEE 802.15.6 MAC reduces the average latency compared to the IEEE 802.15.4 MAC by splitting the superframe slots into two groups containing EAP, RAP, and MAP (optionally with CAP). That is, the IEEE 802.15.6 MAC can allocate the requested slots at two different periods. However, the proposing ASA outperforms that of IEEE 802.15.6 MAC since the ASA can allocate the requested slots according to the application traffic pattern with a fast retransmission mechanism.

**Table 1.** Simulation Properties

Symbol	Description	Value	Unit
$E_F$	Full remaining energy states of slave nodes	$10^4$	joule
$LSA$	Average link speed of slave nodes	250	kbps
$P_{Tx}$	Power consumed in transmitting	52.2	mW
$P_{Rx}$	Power consumed in receiving or listening	56.4	mW
$P_S$	Power consumed in sleeping	3	$\mu$ W
$T_{Sen}$	Average carrier sense time	2	ms
$L_D$	Length of data packet	90	Bytes
$L_R$	Size of ACK/NACK	5	Bytes

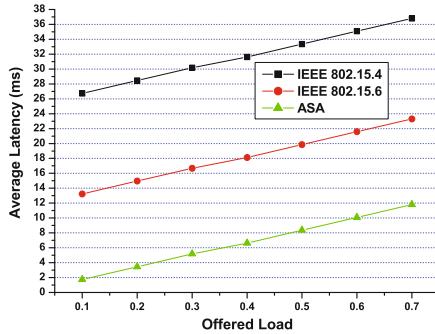
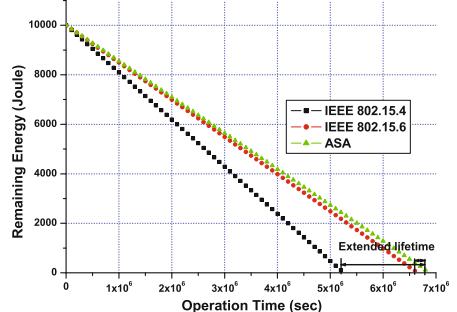
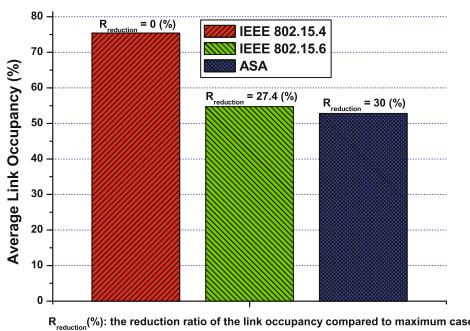
**Fig. 3.** The average latency**Fig. 4.** The remaining energy**Fig. 5.** The average link occupancy of slave nodes (offered load 0.5)

Figure 4 illustrates the remaining energy of slave nodes, where the proposing ASA is superior to the IEEE 802.15.4 MAC and IEEE 802.15.6 MAC. This is because the ASA only wakes up at the reserved transmission time and retransmission time. In the case of the IEEE 802.15.6 MAC enhanced the energy-saving mechanism when compared to the IEEE 802.15.4 MAC, but it still requires a beacon frame in order to know new temporary transmissions at each superframe.

Figure 5 presents the average link occupancy of slave nodes, where the proposing ASA dramatically achieves a 30% link occupancy reduction rate when compared to the maximum case (i.e., the IEEE 802.15.4 MAC) and is superior even to the IEEE 802.15.6 MAC. This is because the proposing ASA minimizes the consumption of link capacity by removing the unnecessary beacon frame exchange and mandatory random access period according to the characteristic of the PMS.

As a result, the proposing ASA thoroughly outperforms both the IEEE 802.15.4 MAC and the IEEE 802.15.6 MAC, with regard to the average latency, the remaining energy, and the average link occupancy in slave nodes.

## 5 Conclusion

The Patient Monitoring System (PMS) is one of the most essential medical systems to monitor the vital signs of a human for 24 hours. Moreover, Wireless Body Area Network (WBAN) can be applied to the PMS. However, existing MAC protocols (i.e., IEEE 802.15.4 MAC and IEEE 802.15.6 MAC) have various limitations to support the time-critical and loss-free requirements of the PMS. Therefore, we proposed an Adaptive Scheduling Algorithm (ASA) to achieve low latency, enhance energy efficiency and attain a low link occupancy for the PMS in WBAN environments. Through the performance evaluation, we proved that the proposed ASA radically outperformed existing IEEE 802.15.4 MAC and IEEE 802.15.6 MAC with regard to average latency, the energy consumption, and the average link occupancy. The proposed ASA will be a helpful reference for the biomedical engineers and medical team staffs, who are looking for a scheduling algorithm to support the energy-efficient, but time-critical characteristics of the PMS.

**Acknowledgment.** This work was supported by the Kyungil University Grant.

## References

1. Shrestha, B., et al.: IEEE 802.15.4 MAC with GTS Transmission for Heterogeneous Devices with Application to Wheelchair Body-Area Sensor Networks. *IEEE Transaction on Information Technology in Biomedicine* (99), 1–11 (2011)
2. <http://www.ieee802.org/15/>
3. Ullah, S., et al.: Performance study of low-power mac protocols for wireless body area networks. In: *IEEE 21st International Symposium on Personal, Indoor and Mobile Radio Communications Workshops*, pp. 112–116 (2010)
4. Faridi, A., et al.: Comprehensive evaluation of the IEEE 802.15.4 mac layer performance with retransmissions. *IEEE Trans. on Vehicular Tech.* 59(8) (October 2010)
5. Tachtatzis, C., et al.: An energy analysis of IEEE 802.15.6 scheduled access modes. In: *IEEE Globecom 2010*, pp. 1270–1275 (2010)
6. Martelli, F., et al.: On the performance of an IEEE 802.15.6 wireless body area network. In: *European Wireless 2011*, pp. 71–76 (2011)

# An Efficient Data Aggregation Scheme in Wireless Sensor Networks

Ying Wang<sup>1</sup> and Guorui Li<sup>2,\*</sup>

<sup>1</sup> Department of Information Engineering, Qinhuangdao Institute of Technology,  
Qinhuangdao, China  
wyqhd@hotmail.com

<sup>2</sup> Electronic Information Department, Northeastern University at Qinhuangdao,  
Qinhuangdao, China  
lgr@mail.neuq.edu.cn

**Abstract.** In wireless sensor networks, the periodically reporting data collection mechanism comes at the cost of power consumption and packet collision. In this paper, we proposed an automatic time series modeling based data aggregation scheme in wireless sensor networks. The main idea behind this scheme is to decrease the number of transmitted data values between sensor nodes and aggregator by using time series prediction model. The proposed scheme can effectively save the precious battery energy of wireless sensor node while keeping the predicted data values of aggregator within application defined error threshold. We show through experiments with real data that the predicted values of our proposed scheme fit the real sensed values very well and fewer messages are transmitted between sensor node and aggregator.

**Keywords:** Wireless sensor networks, Data aggregation, Time series, ARIMA model.

## 1 Introduction

Wireless sensor networks usually consist of many small-sized, low power, inexpensive sensor nodes to monitor some specific phenomenon cooperatively [1]. The communication cost of sensor node is often several orders of magnitude higher than that of computation. For instance, the transmission and reception energy costs for one bit of MICAz node and TelosB node are 600nJ, 670nJ and 720nJ, 810nJ, respectively. However, the computation energy costs for one bit of them are only 3.5nJ and 1.2nJ, respectively. Therefore, data aggregation scheme is often adopted as an effective way to save the precious battery energy of sensor nodes by eliminating the inherent redundancy in the raw data and avoiding unnecessary data transmission.

In this paper, we proposed an automatic ARIMA (Auto Regressive Integrated Moving Average) modeling based data aggregation scheme which utilizes time series model to predict the data of next several periods at both ordinary sensors and aggregators based on the same amount of recent data values. The sensor node will

---

\* Corresponding author.

build an appropriate time series model to predict the future data based on recently sensed values and transmit the parameters of the model to the aggregator automatically. When the prediction error between the sensed value and predicted value is within the application specified error threshold, sensor node will not transmit the sensed value to the aggregator. In this case, the aggregator will regard the predicted value as the sensed value in current data collection period. When the prediction error is beyond the application specified error range, the sensor will rebuild the time series model and transmit the sensed value and new model to the aggregator in order to replace the incorrect predicted value and unsuited prediction model. We show that the predicted values of our scheme fit the real sensed values very well and fewer messages are required to transmit between sensor node and aggregator.

## 2 Related Works

There have been extensive researches in the field of data aggregation scheme in WSNs. Heinzelman et al. proposed LEACH scheme to cluster sensor nodes and let the cluster head to aggregate data in [2]. The cluster head then transmits the aggregated results directly to the sink. Lindsey et al. proposed PEGASIS scheme which organizes all sensors into a chain structure and rotates each node to communicate with the sink [3]. Both LEACH and PEGASIS schemes assume that each node in the network can reach the sink directly in one hop, which limits the size of the network for which they are applicable. Intanagonwiwat et al. proposed GIT scheme which establishes an energy efficient tree by attaching all sensors greedily onto an energy efficient path and prunes less energy efficient paths [4]. However, it might lead to high communication cost in moving event scenarios for the reason of frequently pruning branches. Zhang et al. proposed DCTC scheme which assumes that the distance to the event is known to each sensor and uses the node near the center of the event as the root to construct and maintain the aggregation tree dynamically in [5]. However, it involves heavy message exchanges which might eliminate the benefit of aggregation in large scale networks. Ding et al. proposed EADAT scheme in [6], which is based on energy aware distributed heuristic. It only relies on local knowledge of the network topology and gives higher chances to sensor node with higher residual power to become a non-leaf tree node. Recently, Xu et al. proposed CDA scheme which is based on cooperative communication mechanism. The heuristic algorithm MCT for cooperative data aggregation and its distributed implementation DMCT are also proposed in [7]. However, none of the above schemes have considered the problem of decreasing the number of transmitted data values between ordinary sensors and aggregator. The energy cost of data transmission and reception between them is not trivial. That is the focus and motivation of this paper.

## 3 Automatic ARIMA Modeling Based Data Aggregation Scheme

The automatic ARIMA modeling based data aggregation scheme utilizes ARIMA model to predict the data of next several periods at both ordinary sensors and

aggregators based on the same amount of recently sensed values. They work coordinately to reduce the amount of messages transmitted within the network.

### 3.1 The ARIMA Model

The ARIMA( $p, d, q$ ) model of time series  $\{x_1, x_2, \dots\}$  is defined as

$$\Phi_p(B)\Delta^d x_t = \Theta_q(B)\varepsilon_t \quad (1)$$

$B$  is the backward shift operator,  $\Delta$  is the backward difference,  $d$  is the order of differencing, and  $\Phi_p$  and  $\Theta_q$  are polynomials of order  $p$  and  $q$ , respectively.

$$Bx_y = x_{y-1} \quad (2)$$

$$\Delta = 1 - B \quad (3)$$

ARIMA( $p, d, q$ ) model is the product of an auto regressive part AR( $p$ ):

$$\Phi_p = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (4)$$

an integrating part:

$$I(d) = \Delta^{-d} \quad (5)$$

and a moving average part MA( $q$ ):

$$\Theta_q = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (6)$$

The parameter  $\Phi$  and  $\Theta$  are chosen so that the zeros of both polynomials lie outside the unit circle in order to avoid generating unbounded processes.

### 3.2 Data Aggregation Scheme

The ordinary sensor node runs automatic ARIMA modeling algorithm to build ARIMA prediction model automatically. The notations used in the algorithm are described in Table 1.

**Table 1.** Notations

Notation	Meaning
$\{x_1, x_2, \dots, x_n\}$	Data series
$\{x_1', x_2', \dots, x_n'\}$	Stationary data series
$I$	Differencing order
$\text{diff}(\{x_1, x_2, \dots, x_n\}, I)$	Execute $I$ order of differencing operation to $\{x_1, x_2, \dots, x_n\}$
$\text{variance}()$	Calculate variance
$\varepsilon$	Application defined stationary threshold
$\delta$	Application defined BIC indicator threshold

The automatic ARIMA modeling algorithm works as follows:

```

Collect recently sensed data series { $x_1, x_2, \dots, x_n$ };
I ← 0;
While |variance(diff({ $x_1, x_2, \dots, x_n$ }, I))–variance(diff({ $x_1, x_2, \dots, x_n$ }, I+1))|> $\varepsilon$ 
I ← I+1;
End While
Make { $x_1, x_2, \dots, x_n$ } stationary by  $I$  order differencing and get { $x'_1, x'_2, \dots, x'_n$ };
For AR←1 to MaxAR
  For MA←1 to AR
    Fit ARIMA(AR,0,MA) model according to { $x'_1, x'_2, \dots, x'_n$ } using least
    square method;
    Calculate BIC indicator;
    If (BIC< $\delta$ ) and (Ljung Box white noise test of fit residual passes)
      break ARIMA modeling;
  End If
End For
End For

```

In order to build ARIMA prediction model, sensor node needs to collect recently sensed data series { $x_1, x_2, \dots, x_n$ }. If { $x_1, x_2, \dots, x_n$ } is not stationary, we should make the differencing adjustment to data series until the difference between successive variances is smaller than the application defined stationary threshold  $\varepsilon$ . Then we fit ARIMA prediction model according to the differenced data series { $x'_1, x'_2, \dots, x'_n$ } using least square method. The iteration of ARIMA model fitting process follows the Box search path. It can find an appropriate fitting model using a relatively small number of search times [8]. When the Bayesian Information Criterion indicator of an ARIMA model is smaller than the application defined BIC threshold  $\delta$  and the corresponding Ljung Box white noise test of fit residual passes, the iteration of ARIMA model fitting process will stop. In other words, an appropriate ARIMA prediction model has been built. Here, we choose BIC indicator over AIC indicator for the reason that BIC indicator is more consistent and penalizes free parameters more strongly than AIC indicator.

The automatic ARIMA modeling based data aggregation scheme works as follows:

```

If node is ordinary sensor node
  While (true)
    Run automatic ARIMA modeling algorithm;
    Send ARIMA model parameters to aggregator;
    Do
      predicted value ← predicted data according to ARIMA model;
      If sensed value –predicted value < error threshold
        Historical data [current index] ← predicted value;
      Else
        Historical data [current index] ← sensed value;
        Send sensed value to the aggregator;
    End If

```

```

While |sensed value –predicted value| < error threshold
End While
Else //node is aggregator
    Receive ARIMA model parameters from ordinary sensor;
    While (true)
        Do
            Wait periodical data collection time;
            If received sensed value from ordinary sensor
                Historical data [current index] ← sensed value;
            Else
                predicted value ← predicted data according to ARIMA model;
                Historical data [current index] ←predicted value;
            End If
        Until received ARIMA model parameters from ordinary sensor;
        End While
    End If

```

First of all, the ordinary sensor node runs automatic ARIMA modeling algorithm to build an appropriate ARIMA prediction model. It then sends the ARIMA model parameters to aggregator. After that, it calculates the predicted value according to ARIMA model and compares the sensed value with the predicted value. If the difference between them is less than the predefined error threshold, the sensor node will store the predicted value into historical data queue. Otherwise, it will store the sensed value into historical data queue and send the sensed value to aggregator at the same time. When the predicted value is beyond the fault tolerant range of the sensed value, the ARIMA model will be rebuilt and corresponding ARIMA model parameters of aggregator will be refreshed again.

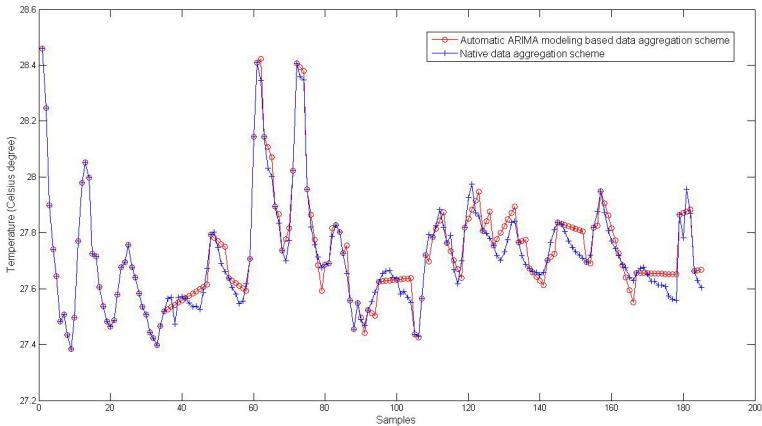
The aggregator listens on the wireless channel to retrieve ARIMA model parameters and sensed values from ordinary sensor node. If the aggregator does not receive any data from sensor node after a predefined periodical data collection time, it means the difference between the sensed value and predicted value is within the acceptable error range. Then the aggregator will calculate the predicted value according to ARIMA model using historical data. Otherwise, it will store the received sensed value into historical data queue and prepare to update the ARIMA model parameters. The periodical data collection time should be selected carefully to ensure it is enough to deliver the message from sensor node to the aggregator. Meanwhile, reliable message retransmission mechanism should be adopted in the underlying MAC layer to guarantee the sensed value could be delivered to aggregator even after collision happens.

## 4 Analysis

In this section, we evaluate and compare the performance of automatic ARIMA modeling based data aggregation scheme with native data aggregation scheme without data prediction. We use the real sensed data collected from TAO (Tropical Atmosphere Ocean) project to demonstrate the performance of our proposed scheme

[9]. The collected data include sea surface temperature, sea level pressure, salinity, relative humidity and density etc along with timestamp information collected once every 10 minutes. We will only use the sea surface temperature data to evaluate our scheme. The other collected measurement will produce the similar results.

In automatic ARIMA modeling based data aggregation scheme, ordinary sensor node will transmit the sensed data value to the aggregator only when the prediction error between sensed value and predicted value is beyond the application specified error threshold. In native data aggregation scheme without data prediction, ordinary sensor node will transmit all the sensed data values to the aggregator. We will refer to it as native data aggregation scheme in the rest of this paper. It's noteworthy that we only consider the problem of data transmission between ordinary sensor node and data aggregator. Both schemes can be combined with other data aggregation schemes which deal with data aggregation between aggregator and sink.

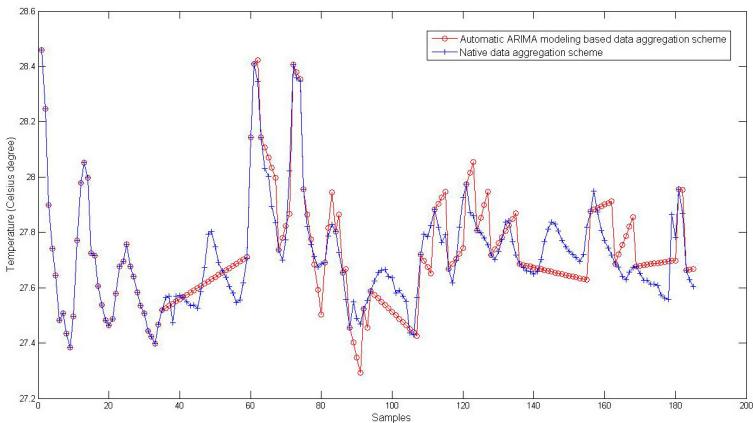


**Fig. 1.** Data comparison of two schemes when the error threshold is set to  $0.1^{\circ}\text{C}$

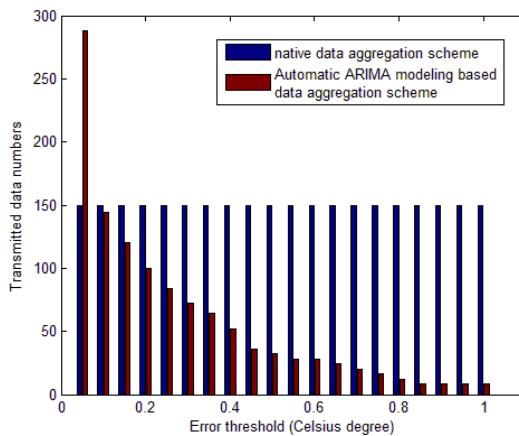
Fig. 1 and Fig. 2 show the comparison of sensed data values of native data aggregation scheme and predicted data values of automatic ARIMA modeling based data aggregation scheme with different predefined error threshold,  $0.1^{\circ}\text{C}$  and  $0.2^{\circ}\text{C}$ , respectively. The source data values which are used to build ARIMA prediction model were collected from the buoy deployed at  $8^{\circ}$  north latitude  $155^{\circ}$  west longitude. We can conclude that the predicted values of our scheme fit the sensed values very well. And the less the predefined error threshold, the better the predicted values fit the sensed values. On the contrary, more ARIMA prediction models should be rebuilt to satisfy the error threshold condition.

Fig. 3 shows the comparison of transmitted data numbers of both data aggregation schemes when the number of predicted values is set to 150. In native data aggregation scheme, all the sensed data values should be sent to the aggregator. In Automatic ARIMA modeling based data aggregation scheme, only the sensed data values which are beyond the error tolerance range and the ARIMA model parameters should be sent to the aggregator. We can see that automatic ARIMA modeling based data aggregation scheme transmits much less number of messages than native data

aggregation scheme for most of the times. Consequently, precious battery energy of wireless sensor nodes are saved and much longer network lifetime is maintained. Only when the error threshold is set too small, many ARIMA prediction models are unfitted and should be rebuilt. Therefore, the transmission of corresponding ARIMA model parameters outnumbers the transmission of sensed data values.



**Fig. 2.** Data comparison of two schemes when the error threshold is set to  $0.2^{\circ}\text{C}$



**Fig. 3.** Comparison of transmitted data numbers

## 5 Conclusion

We have introduced an automatic ARIMA modeling based data aggregation scheme in this paper. Our motivation is to suppress the unnecessary transmitted data values between ordinary sensors and aggregator by data prediction. We first presented the

ARIMA prediction model and then described how the ARIMA prediction model could be built and applied in data aggregation scheme to decrease the number of transmitted messages within the network. Our simulation and analysis indicate that the predicted values of our proposed scheme fit the real sensed values very well and fewer messages are required to transmit between sensor node and aggregator.

**Acknowledgments.** The work is supported by the Fundamental Research Funds for the Central Universities of China (Program No. N100323001), the Natural Science Foundation of Hebei Province, China (Grant No. F2012501014), the Scientific Research Foundation of the Higher Education Institutions of Hebei Province, China (Grant No. Z2010215) and the Science and Technology Project of Liaoning Province, China (Grant No. 2010302005).

## References

1. Yick, J., Mukherjee, B., Ghosal, D.: Wireless Sensor Network Survey. *Computer Networks* 52, 2292–2330 (2008)
2. Heinzelman, W., Chandrakasan, A., Balakrishnan, H.: An Application-specific Protocol Architecture for Wireless Microsensor Networks. *IEEE Transactions on Wireless Communications* 1, 660–670 (2002)
3. Lindsey, S., Raghavendra, C.: PEGASIS: Power-efficient Gathering in Sensor Information Systems. In: 2002 Proceedings of IEEE Aerospace Conference, pp. 1125–1130. IEEE Press, Los Angeles (2002)
4. Intanagonwiwat, C., Estrin, D., Govindan, R., Heidemann, J.: Impact of Network Density on Data Aggregation in Wireless Sensor Networks. In: Proceedings of 22nd International Conference on Distributed Computing Systems, pp. 2–5. IEEE Press, Vienna (2002)
5. Zhang, W., Cao, G.: DCTC: Dynamic Convoy Tree-based Collaboration for Target Tracking in Sensor Networks. *IEEE Transactions on Wireless Communications* 3, 1689–1701 (2004)
6. Ding, M., Cheng, X., Xue, G.: Aggregation Tree Construction in Sensor Networks. In: Proceedings of IEEE 58th Vehicular Technology Conference, pp. 2168–2172. IEEE Press, Piscataway (2003)
7. Xu, H., Huang, L., Zhang, Y., Huang, H., Jiang, S., Liu, G.: Energy-efficient Cooperative Data Aggregation for Wireless Sensor Networks. *Journal of Parallel and Distributed Computing* 70, 953–961 (2010)
8. Yang, S., Wu, Y., Xuan, J.: Time Series Analysis in Engineering Application. HUST press, Wuhan (2007)
9. TAO project, <http://www.pmel.noaa.gov/tao>

# **Design of the Remote Monitoring System for Workshop Based on ZigBee Wireless Sensor Networks**

Hongcheng Yu, Haiping Zhu, Fei He, and Yunlong Wan

State Key Laboratory of Digital Manufacturing Equipment & Technology,  
Huazhong University of Science and Technology, 430074, Wuhan, China  
hongcheng1518@126.com, haipzhu@hust.edu.cn, 26171809@qq.com,  
hust\_wyl@163.com

**Abstract.** To solve the problem of low scalability and poor flexibility in the traditional workshop monitoring system, the designing and implementation methods of the remote monitoring system for workshop based on ZigBee wireless sensor networks were proposed. This paper builds the overall structure of the system, constructs the ZigBee wireless network oriented shop floor, designs the nodes in the net with hardware and software application and uses the information acquainting mode based on B/S structure. This system is capable of achieving the effective remote monitoring for workshop.

**Keywords:** Wireless Sensor Network, Workshop, ZigBee, Remote Monitoring.

## **1 Introduction**

It is the intelligent production and information management that greatly improves productive efficiency and production safety [1][2]. The information collection for manufacturing shop floor is one of the foundations of digital production. Therefore, it is necessary to implement real-time monitoring on the status information of the product, personal, manufacturing resource and complex environment.

The traditional wire shop floor monitoring systems are limited to the problems of high cost, high consumption of power, weak scalability and poor flexibility. With the fast development of the Internet of things, the perception and information exchange between people, objects and environments is affecting people's behavior and thinking habits [3][4]. ZigBee is a standard that defines a set of communication protocols for short-range wireless networking for the applications of low-data-rate transmission between low-power low-cost equipment. The Ad Hoc wireless personal area networks based on ZigBee have some advantages such as large network capacity, good robustness and easy maintenance [5].

Considering the characteristics of ZigBee networks and combining with existed enterprise network, this paper builds a remote monitoring system for workshop, allowing users to achieve ubiquitous real-time monitoring to the shop floor through some terminals such PCs and tablets.

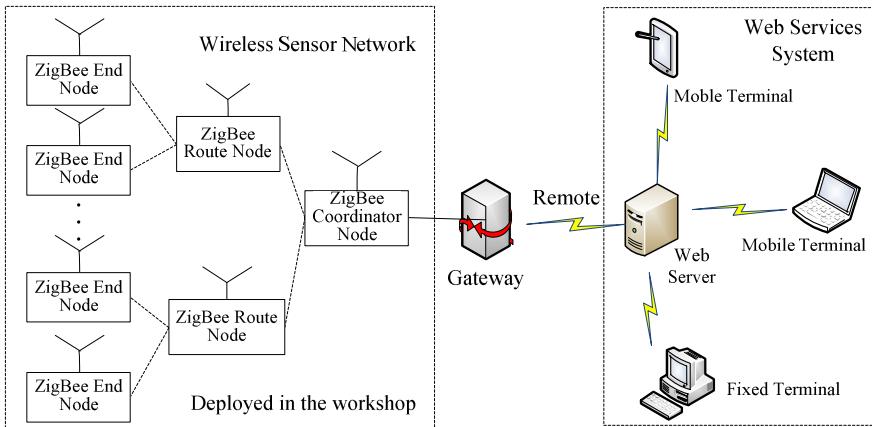
## 2 The Overall Scheme for the System

### 2.1 Requirement for Perceiving Status of Workshop

At an intelligent manufacturing shop floor, we hope to perceive the statuses of production process, personnel and flows of material on instantaneously and accurately to control them effectively. Therefore, it need to collect and process the relevant data in the workshop and consequently, supply the reliable basis for decisions making. The data to be collected can be considered from four aspects, which include the status of equipment, such as equipment's vibration, pressure, temperature, *etc*, the status of personal, such as for duty on time, safe operation, *etc*, flow of material, such as weigh of material, for need on time, *etc*, the environment of the workshop, such as temperature, density of dust. On the other hand, if the staff still is able to access the data of the shop floor in the case without any geographic restrictions, they can achieve more effective real-time monitoring for the status of workshop.

### 2.2 System Structure

In order to satisfy the requirements of intelligent shop floor, we design the remote monitoring system for workshop and the structure is shown in Fig. 1.



**Fig. 1.** Structure of the remote monitoring system for workshop

The system consists of wireless sensor network and Web services system. The wireless sensor network is made up of numerous ZigBee nodes with hierarchical topology deployed in the workshop, which include a ZigBee coordinator node, some ZigBee route nodes and many ZigBee end nodes. Web server is away from the shop floor and belongs to the enterprise data center. Web terminals includes mobile ones, such as Tablet PCs, laptops, *etc*, and fixed ones, for example desktops. Based on the B/S structure, the Web server, together with Web terminals, forms the Web services system, which communicates with the wireless sensor network through the gateway.

The numerous end nodes deployed in the workshop collect the real-time data of the shop floor, and then transmit the data to the coordinator node through route nodes. Through RS232 serial interface cable, the coordinator sends the collected data to the gateway, which will process data and do the protocol conversion to submit the data to the remote Web server. The Web server will parse the received data and store it. Based on these data, the Web serve supplies the Web services, including displaying the real-time data, searching historical data, reminding for warning, setting parameters, *etc.* on the other hand, by means of Web terminals, the users are able to access the Web server to get the related services within the permission in the B/S mode.

### 3 Hardware Design of ZigBee Nodes

The CC2530 produced by TI (Texas Instruments) is a real SoC (System on Chip) solution aiming at the ZigBee applications in the channel of 2.4GHz bands. It can build a powerful network node even with very low material cost. In consideration of these features, such as high sensitivity, strong anti-jamming capability and low power consumption, CC2530 is used as micro-processing module to design the ZigBee nodes in the practical applications with the aid of some kinds of functional circuits.

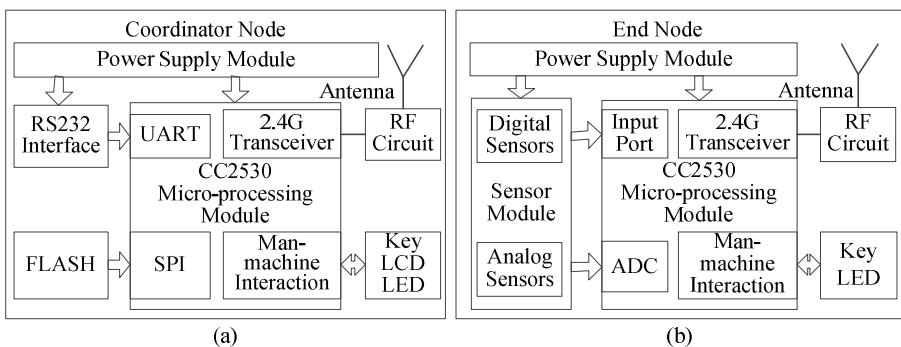


Fig. 2. Block diagrams of ZigBee nodes

**ZigBee Coordinator Node.** The ZigBee coordinator node builds and maintains the wireless sensor work in the workshop and realizes functions of gathering and forwarding data. Besides of CC2530, the main functional circuits include RF (Radio Frequency) module, serial interface, FLASH memory module, power supply module, man-machine interaction, *etc.* The hardware structure is shown in Fig. 2(a).The RF module is used to transmit and receive the radio frequency signal. The communication between coordinator node and gateway is achieved through serial interface. The man-machine interaction can switch the working mode, reset the coordinator node and show the node's information.

**ZigBee Route Node.** The main function of ZigBee route nodes is routing by means of the CC2530 and other functional modules made up of RF (Radio Frequency) modules, power supply module and man-machine interaction.

**ZigBee End Node.** The ZigBee end nodes are the tentacles of the wireless sensor work and supply the raw data for the whole system. Fig. 2(b) shows its hardware structure. Various sensors convert the statuses of the shop floor to electrical signal.

As there are many kinds of monitoring factors in the workshop, it needs to use different kinds of digital and analog sensors. Different digital sensors have different bus connection, such as I2C and SPC and the control pins and data pins are connected to digital I/O ports of CC2530. The signals out of analog sensors will be input to the ADC ports of CC2530 after the adjusting circuit and the IC within the CC2530 will complete the AD conversion of the signal. Physical of ZigBee end node is shown in Fig. 3. The various sensors on the sensor-board connect and communicate with the CC2530 through the sensor interface. At the same time, the sensor-board gets power from motherboard to drive the sensors.

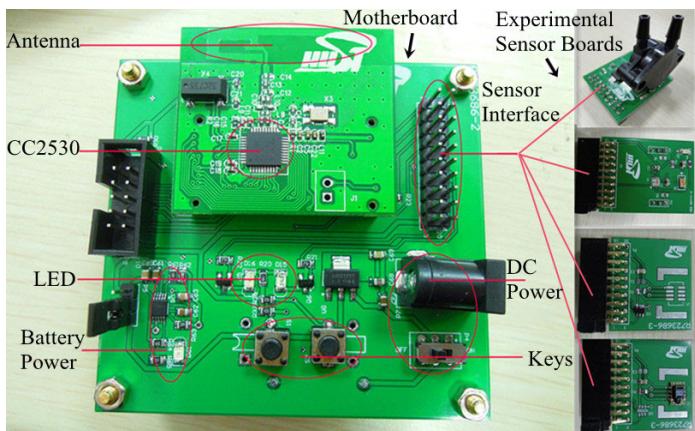


Fig. 3. Actual picture of ZigBee end node

To satisfy the inspection requirements in the workshop, for the monitoring demand in the workshop, the temperature sensors are used to collect the temperature of device operating or environment. Acceleration sensors are used to detect the vibrations of equipment. Proximity sensors will monitor whether the materials arrive on time and the devices are safely operated. Pressure sensors are used to measure the pressure data of devices. And dust sensors are used to detect the working environment in the workshop.

#### 4 Software Design of ZigBee Nodes

Based on the hardware foundation, the software design of each node makes use of the ZigBee protocol stack named Z-Stack supplied by TI. In the whole framework of ZigBee protocol stack, the IEEE802.15.4 defines the bottom two layers, the physical layer and the medium access control layer. Above these, ZigBee standard defines the networking and application layers of the protocol [6]. The functions of the system are achieved mainly in the application layer based on the Z-Stack with trying to call the triggering functions when certain event happens according to the functional requirements.

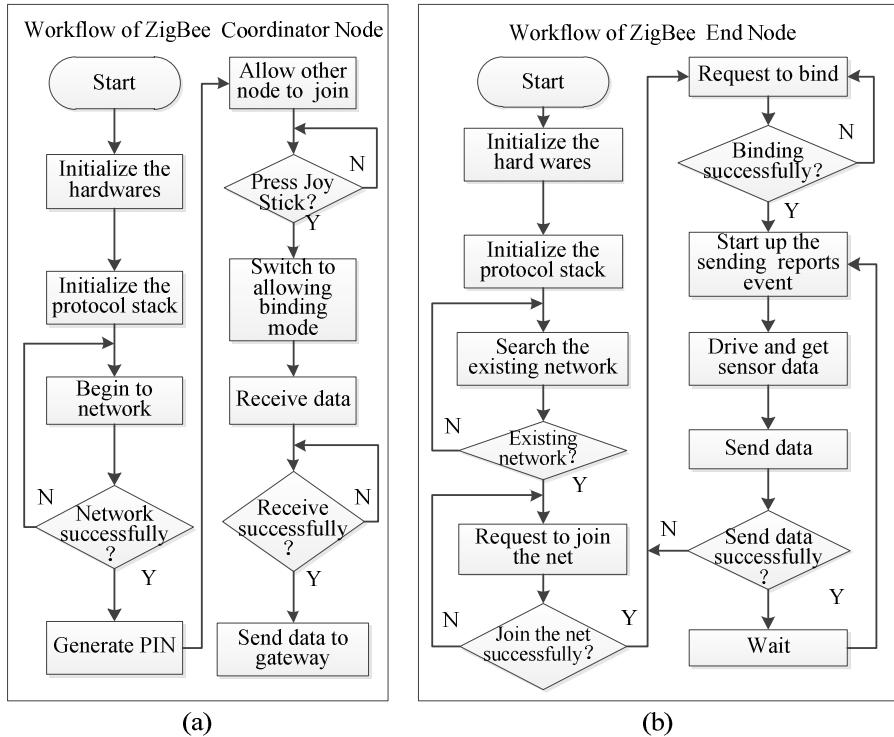


Fig. 4. Workflows of ZigBee nodes

**ZigBee Coordinator Node.** The coordinator node is the center of the whole wireless sensor network in the workshop. It creates, manages, maintains the network and communicates with the gateway. The workflow of ZigBee coordinator node is shown in Fig. 4(a). The coordinator node sends network building broadcasting, broadcasts the PAN ID (Personal Area Network Identity) of the network and sniffs the request for network connection. When it receives the request from other node, the coordinator node judges the request information and then sends response message and allocates network address to the right node if it meets the network conditions. Press the joy stick of the coordinator node to switch it to allowing binding mode and then the end nodes are able to bind and send data to the coordinator node, which parses the received data and sends IEEE address, sensor data and other useful information of the working node to the gateway through the RS232 serial interface.

**ZigBee Route Node.** The main assignments of ZigBee route nodes are to complete the reception and forwarding of data, to ensure the quality of data transmission link and to play the role of the data transmission relay in order to expand the coverage of the wireless sensor network in workshop. After the initiation of the hard wares and protocol stack, the route node tries to join the existed network. In the wireless network, it waits for the sensor data from the end node and transmits it to the coordinator node.

**ZigBee End Node.** The end nodes are responsible for data collection. Its workflow is shown in Fig. 4(b). After the initialization of the node, the end node attempts to join the network and request to bind to coordinator node in the network. Once it is successful bound, the end node will collect data at regular intervals, or sample several times during this period to obtain average, and then send data stored in the custom format packet to the coordinator node.

As its collected object is different, the sensor module of each end node is different from one another. Then modify the APL of the stack of ZigBee and write different data acquisition driver for different sensor, such as driving temperature sensor TMP123 in accordance with SPC-mode, reading data generated by analog sensor like pressure sensor MPX5010 from ADC channels then converting it into the corresponding physical quantity.

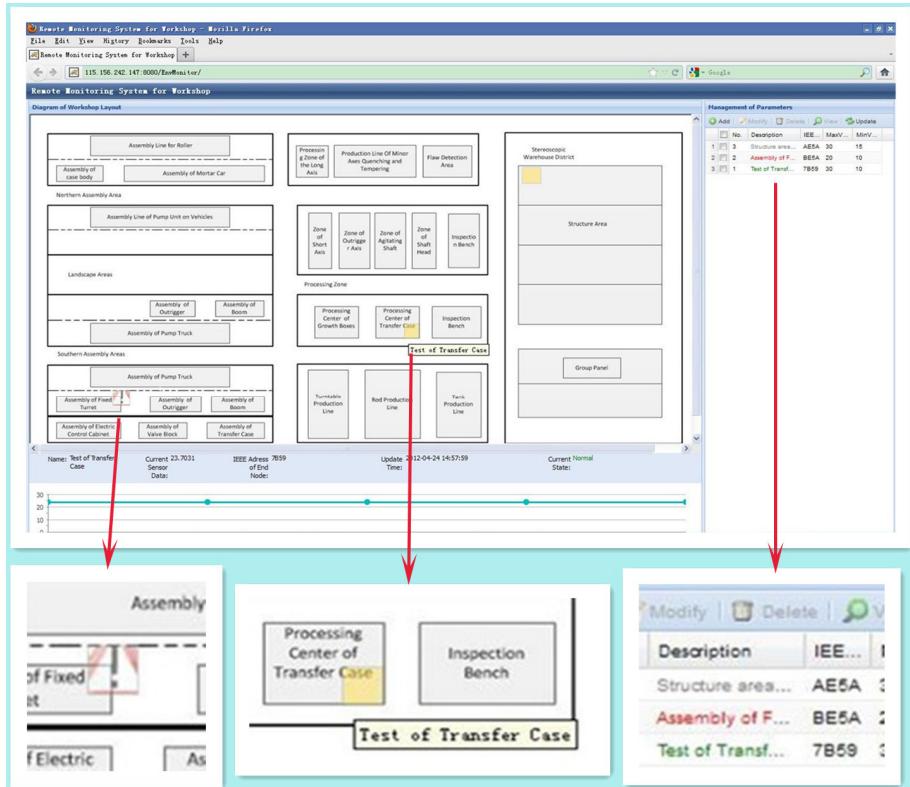
## 5 Remote Monitoring System for Workshop

With respecting to Web services system, we develop the remote monitoring Web services platform for workshop and deploy it on the Web server. The service platform realizes the following two functions.

*Data management.* Serial data receiving program on the gateway is wrote based on VB. The gateway receives the sensor data send from the ZigBee coordinator node through serial interface, makes preliminary data analysis and submits it to the Web server. Then the Web server parses the received data according to the custom format and stores it into the database. The data information received by sever include the receiving date and time, the first eight bits of IEEE addresses of nodes and the sensor data. The system is set to distinguish different end nodes by their first eight bits of IEEE addresses whereby to get information of nodes from the database.

*Remote information services.* On the basis of data management, the platform supplies functions such as display of real-time data, querying of history data and setting of parameters of remote nodes. Users are able to browse related information and manage nodes. The interface of user terminal is shown in Fig. 5.

The main part of the interface is the layout of workshop, and the yellow boxes represent the locations of the ZigBee end nodes in the shop floor. When focusing on certain terminal node, it show the real-time data collected by this node on the bottom of the interface. Fig. 5 shows the end node named “Test of Transfer Case” deployed in the middle is on the status of normal, the one on the lower left shown as an exclamation mark in the red triangle box is out of limit and the upper right one is offline for too long time. At the same time, the corresponding descriptions on the panel named “Management of Parameters” are shown respectively as the colors of green, red and gray. Double-click the yellow box, it will show the history data of the corresponding end node. On the panel “Management of Parameters”, the users are able to manage the information of end nodes by operations of addition, modification and deletion.



**Fig. 5.** Interface of user terminal

## 6 Conclusion

In order to obtain the real-time status information of the workshop accurately, this paper introduces ZigBee wireless communication technology into the application of collecting the data in the shop floor and designs the remote monitoring system for workshop based on ZigBee wireless sensor networks combined with existed enterprise network applications. It has the following characteristics:

- (1)With the use of the ZigBee wireless sensor network to collect and transmit the data, it changes the traditional wired monitoring way and builds a system with low cost, flexibility and real-time.
- (2)It allows the terminals, especially the mobile ones, to get remote sensor data flexibly in Web mode.
- (3)The system has good scalability. By adding more types of sensors, designing the adjusting circuit and connecting to the motherboard of ZigBee end node through sensor interface, it will meet more requirements for the workshop condition monitoring.

These urgent-needed functions have been successfully realized. Based on the function of data displaying and false alarming, some other advanced functions like intelligent decision and security policy would probably be the research focuses.

**Acknowledgements.** The authors greatly acknowledge the financial supports from the Pre-research Foundation of General Armament Department of China with the Grant number 51318010104, the National Key Technology R&D Program of China with the Grant number 2012BAF10B08.

## References

1. Bi, K., Gao, W.: Measuring and Method of Informatization Level for Manufacturing Enterprise. In: 2011 International Conference on Information Science and Technology (ICIST), pp. 520–527. IEEE Press, Nanjing (2011)
2. Niu, Z.-W., Fu, J.-N., Guo, W., Qi, E.-S.: Manufacturing Informatization Development Strategy Based on System Dynamics. Journal of Tianjin University (Social Sciences) 11(2), 106–109 (2009) (in Chinese)
3. Kopetz, H.: Real-Time Systems: Design Principles for Distributed Embedded Applications. Springer US, New York (2011)
4. Atzoria, L., Jerab, A., Morabitoc, G.: The Internet of Things: A survey. Computer Networks 54(15), 2787–2805 (2010)
5. Sung, W., Hsu, Y.: Designing an Industrial Real-time Measurement and Monitoring System Based on Embedded System and ZigBee. Expert Systems with Applications 38(4), 4522–4529 (2011)
6. Farahani, S.: ZigBee Wireless Networks and Transceivers. Newnes, Burlington (2008)

# **Delay Aware Time Slot Allocation Algorithm in Light-Trail Network**

Minglei Fu, Yiluan Zhuang, Maolin He, and Zichun Le<sup>\*</sup>

College of Science, Zhejiang University of Technology, Liuhe Road.288,  
310023 Hangzhou, China

{fuml, lzc}@zjut.edu.cn, yiluan\_z@yahoo.cn, 421695729@qq.com

**Abstract.** Light trail (LT) is a generalization of lightpath with ability to provision emerging IT applications such as data center, cloud computing, etc. The time slot allocation (TSA) problem is addressed with respect to delay aware ability. In addition, a delay aware TSA algorithm is given on the basis of bidding policy. The bid value is composed of two parts: the delay bid and the buffer bid. The node which sends the highest bid will get the transmission right of the light trail for the next time slot. Simulation results concerning the traffic and delay in a 5 and 6 nodes light trail network show that: by means of adopting the delay aware TSA algorithm, higher loads which usually need longer delay are allocated the time slots preferentially. Hence, the delay aware TSA algorithm can improve the LT network throughput when assures the transmitting delay less than the maximum-tolerant delay.

**Keywords:** time slot allocation, light trail, delay aware.

## **1 Introduction**

Light-trails (LT) have been proposed as a generalization of a lightpath and are efficient high-speed networking solutions to provide dynamic bandwidth allocation, optical layer multicasting, sub-wavelength spatial grooming at low price-points using contemporary technology [1]. In addition, connection provisioning in light-trail does not require nodal reconfiguration, implying no optical layer switching and thereby is most suitable for dynamic scenarios. Hence, those remarkably advantages of LT makes it can be adopted by the basic infrastructure for data center, cloud computing, enterprise applications, etc [1-2].

LT network is a time-slotted system in nature, especially on the data plane. Time slot allocation (TSA) which also called dynamic bandwidth allocation or scheduling is one of the key algorithms (or protocols) which deeply affect the performance of LT network. Although the dynamic bandwidth allocation (DBA) problems have been addressed extensively on the other optical networks, such as passive optical network (PON) based networks, the unique features of LT network makes it still attractive to solve the TSA problem. In this paper, we focus our attention on the delay aware problem in TSA and propose a delay aware TSA algorithm.

---

<sup>\*</sup> Corresponding author.

The paper is organized as follows: In section 2, related works on TSA are summarized. In section 3, we discuss the delay aware problem of TSA in LT network and section 4 gives the proposed algorithm. Section 5 shows the simulation results of the algorithm and the paper is concluded in section 6.

## 2 Related Work

In this section, the most relevant works on the TSA schemes in LT network were summarized. Among them, Dual auction opportunistic protocol (DAOP) and Delay sensitive smoothed round robin (DS2R2) were considered as the two main contributions to the TSA problem.

The DAOP consisted of two stages-bidding and assignment [3]. Each node periodically placed a bid for a light-trail depending on network parameters and its bandwidth requirements. The advantage of the DAOP was that the bidding and assignment were done ahead in time as compared to the establishment of the connection-hence maximizing efficiency. In [4], the DAOP was formally stated, analyzed using Markov models. And simulation showed that it resulted in 45% betterment over existing linear program (LP) or heuristic models. An adaptive round time (ART) MAC protocol for closed light-trail (CLT) was presented [5]. With the round time adapted to the current load, ART gains high wavelength utilization while the shared bandwidth was guaranteed for each node by a pre-assigned maximum transmission window. Further, the performance of LT-FA MAC and ART MAC were evaluated [6]. In addition, a dynamic bandwidth allocation scheme named Demand and Delay latency-aware with Two-round Evaluation (DDTE) was proposed on the basis of DAOP [7]. The length of a time slot was variable in DDTE while Dual Auction used time slots with a fixed length. And DDTE deal with time slot allocation to multiple nodes in a cycle. In [8], DAOP was adopted by the superscheduler architecture and job migration algorithm for computational grids over LT WDM networks.

DS2R2 was used for bandwidth provisioning in LT networks [9]. The protocol guaranteed efficient scheduling and took into consideration both bandwidths as well as delay profile of the participating users. In [10], DS2R2 was used as the static centralized control scheme on the control plane for service provisioning in LT WDM optical ring networks.

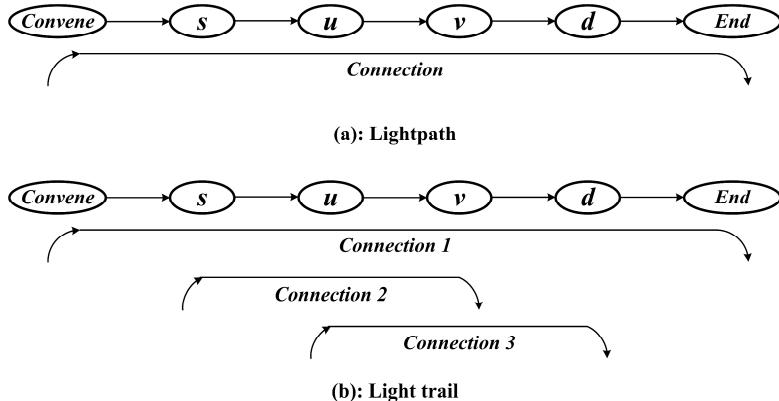
Although delay had been considered as one of the key factors in both DAOP and DS2R2, it was not addressed thoroughly. And, in this paper, we would like to discuss the delay aware TSA problem.

## 3 Delay Aware in TSA of LT Network

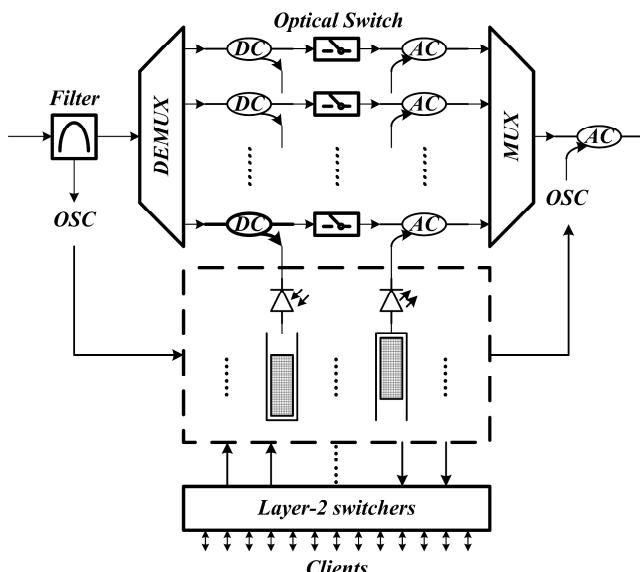
In this section, we introduce the concept of light trail and show the difference between lightpath and light trail. In addition, a typical LT node structure is given. Then, we describe how the TSA is implemented in the LT connections. At last, we address the delay aware in the LT network.

### 3.1 LT Network and Its Typical Node Structure

Light-trail is a generalization of a lightpath. A lightpath is a point-to-point optical circuit, while a light-trail is a multipoint-to-multipoint unidirectional wavelength bus [1-2]. As shown in Fig.1, only one connection is allowed to be built via one wavelength from the convene node to the end node. However, multiple nodes can take part in communication along the light-trail which results in more connections being established. And Connections are provisioned over the wavelength bus without any optical switch configuration.



**Fig. 1.** Comparison of lightpath and light trail

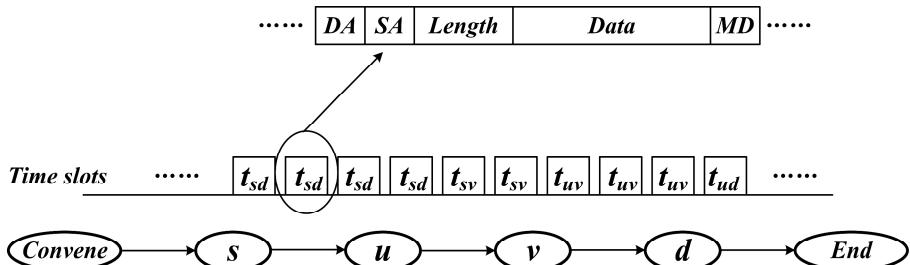


**Fig. 2.** Typical node structure of LT network

To facilitate a light-trail, every node must be able to support the drop-and-continue, passive add, and burst-mode optics features in the optical layer [1]. The typical structure of LT node is composed of optical filter, demultiplexer, drop coupler (DC), optical switch, add coupler (AC) and multiplexer, as shown in Fig.2. Among these, the 3dB coupler is usually adopted as the AC, while optical splitters are chosen as the DC, whose the splitting ratio are 90/10, 80/20 and 70/30 in sequence on the LT test-bed. Operating principle of the typical LT node can be described as follows: The optical service channel (OSC) wavelength which carries out-of-band (OOB) control signaling is filtered by the optical filter and then be sent to the LT control unit. Rests of the multiplexed data wavelengths are demultiplexed and pass the DCs and ACs. DCs and ACs are used to drop or add the data wavelength according to the information obtained by the OSC signals. Next, rests of the bypass data wavelengths and the newly added data wavelengths are multiplexed by the multiplexer. At last, the newly generated OSC wavelength is added by the AC and joins the multiplexed date wavelengths. Optical switches are used to establish or cut down the light-trails.

### 3.2 TSA in LT Network

Once a light-trail is established between the convene node and end node, nodes except the end node send their requests to the end node through the OOB control channel and then the end node runs the TSA algorithm (protocol) to determine which node should transmit the traffic in the next time slots. Usually, the slot sizes are comparable with an integral multiple of the SONET slot (125ms) [1].



**Fig. 3.** Time slots allocation during the LT communication

As shown in Fig.3, a six node light-trail has been established between the convene node and end node. Node s, u, v, d are the intermediate nodes.  $tsd$  denotes the time slot allocated for the connection from s to d. Among the frame structure of time slot, we pay attention to the destination address (DA), source address (SA), length (data size), data and maximum-tolerant delay (MD).

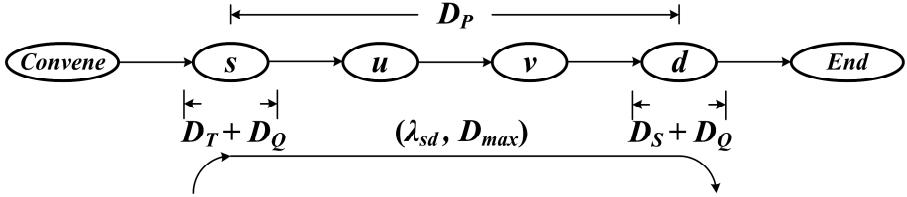
The TSA problem in the LT network can be stated as follows: For a set of requests, allocate the time slot for the connection from the source address to the destination address to meet the bandwidth requirement subject to the constraints that all the traffic must be transmitted before its delay reaches the maximum-tolerant delay.

### 3.3 Delay Aware in LT Network

In the LT network, delay during the connection between two nodes can be divided into four parts: 1) Transmission delay (DT); 2) Queuing delay (DQ); 3) Propagation delay (DP); 4) Slot-synchronization delay (DS). As shown in Fig.4, a connection has been built to transmit the traffic load (denoted by  $\lambda_{sd}$ ) with the maximum-tolerant delay (denoted by  $D_{max}$ ). At node s, DT is caused by transmitting the data and DQ is caused by the queue behavior. DP is caused by propagation progress. And, at node d, DS is caused by synchronizing to the time slots for communication with its neighbor. Hence, let  $D_{Total}$  denote the total delay, then:

$$D_{Total} = D_T + 2D_Q + D_P + D_S. \quad (1)$$

Note that  $D_{Total}$  must less than the  $D_{max}$ .



**Fig. 4.** Delay of the connection from  $s$  to  $d$

Let  $1/\mu$  denote the average packet size in the slots,  $1/\rho$  denote the slot size and  $C_{sd}$  denote the assigned link capacity between node  $s$  and  $d$ . Let  $L$  denote the fiber length between node  $s$  and  $d$ . Let  $C_t$  denote the propagation speed over the fiber. Then,

$$D_T = 1/\rho C_{sd}. \quad (2)$$

$$D_Q = \lambda_{sd}/(\mu C_{sd}(\mu C_{sd} - \lambda_{sd})) [11]. \quad (3)$$

$$D_S = 1/(2\rho C_{sd}) [11]. \quad (4)$$

$$D_P = L/C_t. \quad (5)$$

Usually, DP is negligible when compared with DT, DQ and DS . Hence, the delay of connection from node  $s$  to  $d$  can be approximated by equation (6):

$$D_{Total} = 2\lambda_{sd}/(\mu C_{sd}(\mu C_{sd} - \lambda_{sd})) + 3/(2\rho C_{sd}) \quad (6)$$

## 4 Delay Aware TSA Algorithm

### 4.1 Basic Idea

As stated above, the required bandwidth and maximum-tolerant delay of the requests are two essential parameters for the end to determine which node can get the right to transmit the traffic in the next time slot. In the delay aware TSA algorithm proposed

in this paper, each node except the end will compute and place a bid for the next time slot at the end of the current time slot. And the bid value is composed of two parts: the delay bid and buffer bid. Then, the end node will receive all the bids from the node in the light trail and compute the highest bid. The node which sends the highest bid will get the transmission right of the light trail for the next time slot. At last, the end node will inform all the nodes in the light trail of the successful node by the OOB channel and the other nodes can compute and send their new bids in the next round.

## 4.2 Algorithm Procedures

Firstly, some notations used in the algorithms are given.

- N: Number of nodes which intend to send data via the LT,  $N_i$  a particular node;
- $B_i$ : The value of buffer size at  $N_i$ ;
- $B_{MAX_i}$ : The maximum buffer size at  $N_i$ ;
- $B_{Ratio_i}$ :  $B_i / B_{MAX_i}$ ;
- M: Number of types of traffic,  $M_i$  a particular traffic;
- $D_{Total}(M_i)$ : Delay of  $M_i$  from  $N_i$  to the destination node;
- $D_{MAX}(M_i)$ : Maximum-tolerant delay of  $M_i$ ;
- $D_{Ratio_i}$ :  $D_{Total}(M_i) / D_{MAX}(M_i)$ ;
- $BID(N_i, M_i)$ : Bid value of  $N_i$  for  $M_i$ ,  $BID(N_i, M_i) = \max\{B_{Ratio_i}, D_{Ratio_i}\}$ ;

Procedures of the delay aware TSA algorithm can be stated as follows:

- Step 1:** For each  $M_i$ ,  $N_i$  computes all the  $D_{Total}(M_i)$  according to Equation (1);
- Step 2:** For each  $M_i$ ,  $N_i$  computes  $D_{Ratio_i}$ , and choose the  $M_k$  with highest  $D_{Ratio_i}$ ;
- Step 3:**  $N_i$  computes  $B_{Ratio_i}$  and  $BID(N_i, M_k)$ , send the  $BID(N_i, M_k)$  to end node;
- Step 4:** For all the nodes, end node chooses the highest  $BID(N_m, M_n)$ .
- Step 5:**  $M_n$  at  $N_m$  gets the transmission right of the LT during the next time slot.
- Step 6:** The other nodes update the  $D_{Total}(M_i)$  by adding  $D_{Total}(M_n)$ , goto Step 1.

## 5 Simulation Results

We simulate a unidirectional light-trail with node numbers from 4 to 6. Parameters used in the simulation are: the link capacity is 10Gbps, buffer size of each node is 10Mbyte, maximum-tolerant delay is 3.0ms, packet size is 1kbit and slot size is 1.036Mbit. During the simulation, the traffic matrix follows a Poisson distribution with mean value is 5Mbps.

Fig.5 and fig.6 show the simulation results in terms of traffic (or real-time throughput) and average transmitting delay when the node number of LT is 5. In fig.5, we also give the scheduled node sequence when transmitting the time slots. And fig.6 shows that the maximum delay is 2.0364ms, which less than the pre-set maximum-tolerant delay.

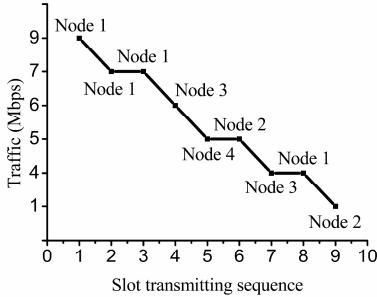
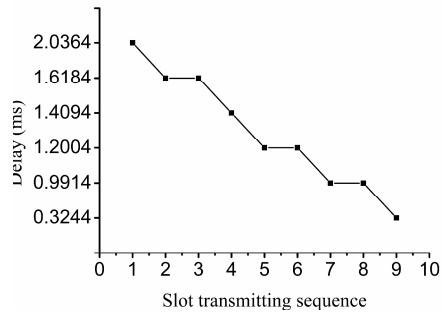
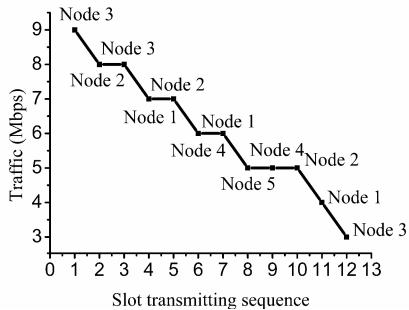
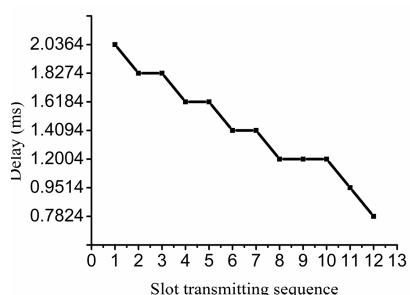
**Fig. 5.** Traffic of LT with 5 nodes**Fig. 6.** Delay of LT with 5 nodes

Fig.7 and fig.8 show the simulation results in terms of traffic (or real-time throughput) and average transmitting delay when the node number of LT is 6. In fig.7, we also give the scheduled node sequence when transmitting the time slots.

**Fig. 7.** Traffic of LT with 6 nodes**Fig. 8.** Delay of LT with 6 nodes

Simulation results from fig.5 to fig.8 show that: by means of adopting the delay aware TSA algorithm, higher loads which usually need longer delay are allocated the time slots preferentially. Hence, the delay aware TSA algorithm can improve the LT network throughput when assures the transmitting delay less than the maximum-tolerant delay.

## 6 Conclusion

In this paper, we address the time slot allocation problem in light trail network and a novel delay aware TSA algorithm are proposed. Simulation results show that the proposed algorithm aims at allocating the time slot for the highest load in each bid round, which improves the LT network traffic. Besides, the average transmitting delays are less than the maximum-tolerant delay.

**Acknowledgments.** This work was supported by Natural Science Foundation of China (No.61172081), Natural Science Foundation of Zhejiang Province (No.LQ12F05008), Public Welfare Technology Application Research Plan of Zhejiang Province (No.2011C21011) and Natural Science Foundation of Zhejiang University of Technology (No.2011XY027).

## References

1. Somani, A., Gumaste, A.: Light-trails: Distributed optical grooming for emerging data-center, cloud computing, and enterprise applications. In: Opt. Fiber Commun. Conf. Expo. Natl. Fiber Opt. Eng. Conf., OFC/NFOEC (2011)
2. Gokhale, P., Kumar, R., Das, T., Gumaste, A.: Cloud computing over metropolitan area WDM networks: The light-trails approach. In: GLOBECOM IEEE Global Telecommun. Conf. (2010)
3. Gumaste, A., Zheng, S.Q.: Dual auction (and recourse) opportunistic protocol for light-trail network design. In: Int. Conf. Wireless Optic. Com. Netw. (2006)
4. Gumaste, A., Das, T., Mathew, A., Somani, A.: An autonomic virtual topology design and two-stage scheduling algorithm for light-trail WDM networks. J. of Opt. Comm. and Netw. 3, 372–389 (2011)
5. Tran, N., Vu, N., Tran, D., Park, J.: ART: A dynamic medium access protocol for closed light-trail networks. In: HUT-ICCE - Int. Conf. Commun. Electron, pp. 144–149 (2008)
6. Fukushima, Y., Tanaka, K., Yokohira, T.: Performance evaluation of medium access control methods in light trail networks. In: Int. Conf. Adv. Commun. Technol. ICACT, vol. 2, pp. 1421–1425 (2009)
7. Hsu, C.F., Hsu, K.K., Ku, C.H.: An efficient dynamic bandwidth allocation algorithm with two-round deliberation in light-trail networks. In: Proc. Int. Conf. Netw.-Based Inf. Syst., NbiS, pp. 260–264 (2010)
8. Gumaste, A., Jain, S., Somani, A.K.: An efficient superscheduler architecture and job migration algorithm for computational grids over light-trail WDM networks. In: Proc. Int. Conf. Broadband Commun., Netw. Syst., BROADNETS (2009)
9. Bafna, P., Gumaste, A., Ghani, N.: Delay sensitive smoothed round robin (DS2R2) scheduler for high-speed optical Networks. IEEE Commun. Lett. 11, 628–630 (2007)
10. Gumaste, A., Chandarana, J., Bafna, P., Ghani, N., Sharma, V.: On control plane for service provisioning in light-trail WDM optical ring networks. In: IEEE Int. Conf. Commun., pp. 2442–2449 (2007)
11. Reaz, A., Ramamurthy, V., Sarkar, S., Ghosal, D., Dixit, S., Mukherjee, B.: CaDAR: An efficient routing algorithm for wireless-optical broadband access network. In: IEEE Int. Conf. Commun., pp. 5191–5195 (2008)

# Quality of Recovery (QoR) Analysis for the ONU Protection in WOBAN

Zichun Le, Maolin He, Yiluan Zhuang, and Minglei Fu<sup>\*</sup>

College of Science, Zhejiang University of Technology, Liuhe Road.288,  
310023 Hangzhou, China

{lzc,fuml}@zjut.edu.cn, 421695729@qq.com, yiluan\_z@yahoo.cn

**Abstract.** The wireless-optical broadband access network (WOBAN) is a promising architecture for future access networks. In this paper, the survivability problem of WOBAN with respect to ONU protection is addressed. And, a new concept called Quality of recovery (QoR) is introduced as the assessment of different protection schemes. Availability, recovery time, redundancy are chosen as the main factors in QoR. To verify the analytical models for four ONU protection schemes, simulation has been done on the basis of a WOBAN with 32 ONUs. Simulation results show that the weakest selection protection scheme is seemed as the scheme with best cost effective when compared with other schemes.

**Keywords:** quality of recovery, WOBAN, ONU, protection.

## 1 Introduction

A wireless-optical broadband access network (WOBAN) is a combination of wireless and optical network segments to optimize the cost and performance of an access network. Due to the heterogeneous network environments of WOBAN, network survivability has become one of most important issues when the WOBAN is deployed. Some works concerning the protection schemes for WOBAN have been reported. In [1], a cost-effective protection for WOBAN that deals with network element failures in the optical part was proposed. A probabilistic analysis of the survivability of NG-PONs and hybrid fiber-wireless access networks was proposed in [2]. The problem of planning a fault-tolerant multiradio WOBAN while using resources efficiently was proposed in [3].

However, rarely work has been reported on how to assess these protection schemes with a unified aspect. Recently, the concept of quality of recovery (QoR) has been proposed [4-9]. By adopting QoR, a quality assessment and differentiation of the recovery methods for the network resilience can be realized. Hence, in this paper, the QoR is introduced to assess different ONU protection schemes in WOBAN.

---

\* Corresponding author.

## 2 ONU Protection Schemes in WOBAN

A typical WOBAN architecture is composed of optical line terminal (OLT), splitter, optical network unit (ONU) and fibers in the optical part and gateway, wireless router in the wireless part. Different protection schemes are focused on ONUs to provide survivability against fiber link failures [2]. Among them, two types of ONU protection schemes are interesting: (1) ONU with wireless protection; (2) ONU with both optical and wireless protection. In addition, the first scheme can be furtherly divided into three methods: ①Random selection; ②Weakest selection; ③Strongest selection.

The ONUs with wireless protection are able to communicate with each others in wireless manner when the fiber links from ONU to OLT are failed. Let  $m$  denote the number of ONUs in M. Let  $n$  denote the number of ONUs in WOBAN. Then, the three protection methods belong to scheme 1 can be stated as follows:

- ① Random selection: M are randomly selected among  $n$  ONUs.
- ② Weakest selection: The  $m$  ONUs with smallest probability of being (optically) connected to the OLT are selected.
- ③ Strongest selection: Conversely, M is selected with the largest probability of being (optically) connected to OLT.

ONU with both optical and wireless protection combines the optical and wireless protection together by connecting  $ONU_1$  with additional backup fiber linking to OLT, and then upgrading the  $m-1$  weakest ONUs (smallest well-connected probability) with MPPs.

## 3 ONU Protection Schemes on the Basis of QoR

In this section, a particular assessment model based on QoR was presented for different ONU protection schemes (which are defined specifically in section 2). For describing the QoR of different protection schemes, three proper factors, availability, recovery time, redundancy, are chosen. In the followings, we discuss the QoR model around these factors as the sequence of Abstraction, Normalization and Application.

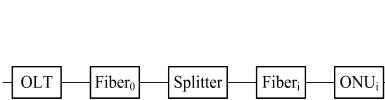
### 3.1 Abstraction

This procedure is related to the calculation or estimation of some real parameter. Now,  $A$  (availability),  $T$  (recovery time),  $R$  (redundancy) will be calculated in specific.

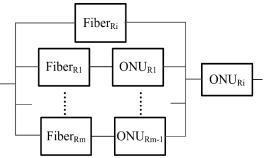
#### 3.1.1 Availability

The factor, availability, means the risk of the connection or component, e.g.,  $A_{OLT}$  is the probability that OLT doesn't work properly, and  $A_{Fiber}$  is the risk of disconnection of the fiber. For a protection scheme, availability is the robustness of the protection

network. The calculation of the availability is based on the reliability block diagram (RBD) of the protection network. Here, in every protection scheme, due to the states of ONUs are different (some ONUs are selected to achieve a high available, but the others are not), we choose a certain unit,  $ONU_i$ , to analyze the availability of it. Then calculate the whole  $A$  of the protection scheme.



**Fig. 1.** RBD of the basic architecture with non-protection



**Fig. 2.** RBD of the Random selected protection scheme

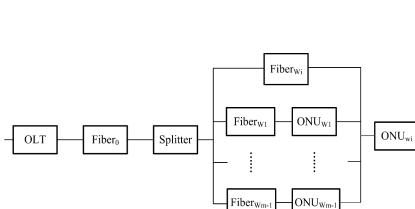
In Fig.1, the availability  $ONU_i$  of the basic architecture can be express as:

$$A_{B-ONU_i} = A_{OLT} \times A_{Fiber_0} \times A_{Splitter} \times A_{Fiber_i} \times A_{ONU_i} \quad (1)$$

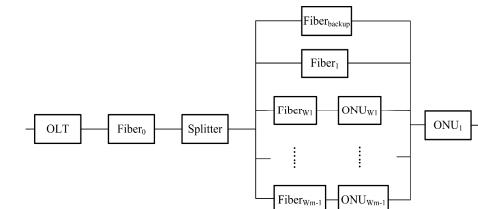
For the architecture of Random selection protection scheme shown in Fig.2, the availability of  $ONU_i$  which is luckily selected is:

$$A_{R-ONU_i} = A_{OLT} \times A_{Fiber_0} \times A_{Splitter} \times \left[ 1 - \left( 1 - A_{Fiber_i} \right) \times \prod_{j \in M, j \neq i} \left( 1 - A_{Fiber_j} \times A_{ONU_j} \right) \right] \times A_{ONU_i} \quad (2)$$

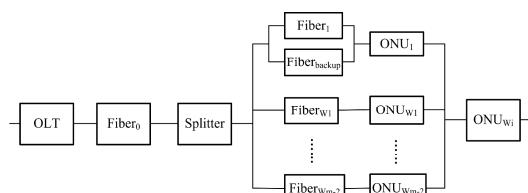
Where set  $M$  is random selection of all ONUs. Obviously, if  $ONU_i$  is not in set  $M$ , the availability equals  $A_{B-ONU_i}$ .



**Fig. 3.** RBD of the Weakest selected protection scheme



**Fig. 4.** RBG of the Optical and wireless protection scheme for  $ONU_1$



**Fig. 5.** RBG of the Optical and wireless protection scheme for  $ONU_{i \in M}$

In Fig.3, the availability of selected ONU<sub>i</sub> related to the architecture is express as:

$$A_{W-ONU_i} = A_{OLT} \times A_{Fiber_0} \times A_{Splitter} \times \left[ 1 - (1 - A_{Fiber_i}) \times \prod_{j \in M, j \neq i} (1 - A_{Fiber_j} \times A_{ONU_j}) \right] \times A_{ONU_i} \quad (3)$$

Where, set M is defined by Weakest selection scheme. And the availability of ONU which is not in set M is  $A_{B-ONUi}$ .

The availability  $A_S$  of Strongest selection protection scheme is similar to the Weakest scheme, but defines set M by Strongest selection scheme. The calculations of  $A_{O\&W-ONUi}$  in the Optical and wireless protection scheme have three cases, as shown in Fig.4 and Fig.5, the formula is as follows:

$$A_{O\&W-ONU_1} = A_{OLT} \times A_{Fiber_0} \times A_{Splitter} \times \left[ 1 - (1 - A_{Fiber_{backup}}) \times (1 - A_{Fiber_1}) \times \prod_{j \in M} (1 - A_{Fiber_j} \times A_{ONU_j}) \right] \times A_{ONU_1} \quad (4)$$

$$A_{O\&W-ONU_i} = A_{OLT} \times A_{Fiber_0} \times A_{Splitter} \times \left\{ 1 - [1 - A_{ONU_i} \times (A_{Fiber_i} + A_{Fiber_{back-up}} - A_{Fiber_i} \times A_{Fiber_{back-up}})] \times \prod_{j \in M} (1 - A_{Fiber_j} \times A_{ONU_j}) \right\} \quad (5)$$

In formula (4),  $i=1$  and the set M is  $m-1$  ONUs with Weakest selection.

In formula (5),  $i \notin M, i \neq 1$  set M is  $m-2$  ONUs with Weakest selection. For others, the availability of ONU is  $A_{B-ONUi}$ .

Now, the whole A of the basis architecture and architectures of four protection schemes and are calculated as follows:

$$A_B = \sum_{i=1}^n A_{B-ONU_i} \quad (6)$$

$$A_R = \sum_{i=1}^m A_{R-ONU_i} + \sum_{i=1}^{n-m} A_{B-ONU_i} \quad (7)$$

$$A_W = \sum_{i=1}^m A_{W-ONU_i} + \sum_{i=1}^{n-m} A_{B-ONU_i} \quad (8)$$

$$A_S = \sum_{i=1}^m A_{S-ONU_i} + \sum_{i=1}^{n-m} A_{B-ONU_i} \quad (9)$$

$$A_{O\&W} = A_{O\&W-ONU_1} + \sum_{i=1}^{m-2} A_{O\&W-ONU_i} + \sum_{i=1}^{n-m} A_{B-ONU_i} \quad (10)$$

### 3.1.2 Recovery Time

In a protection scheme, another important factor is the time cost in the recovery. The recovery time includes, the failure detection time ( $T_d$ ), the notification time ( $T_n$ ), switching time ( $T_s$ ) for optical protection and calculation time ( $T_c$ ) of new path in the wireless protection. When the structure has no protection at all, the recovery time equals the time of fixing the failure components, in structure presented in our paper, the formula is:  $T_B = n \times \sum (1 - A_i) \times T_i$ , where  $A_i$  is the availability of the component and  $T_i$  is the repairing time of a component.

In the case of the wireless protection scheme, which includes Random Selection, Strongest Selection and Weakest Selection, the recovery time equals the result of adding two parts, the selected ONUs and the not selected ones:  $T_{Wireless} = m \times (T_d + T_n + T_c) + (n - m) \times T_B$ . For the Optical and wireless protection scheme, the formula can be expressed as:

$$T_{O\&W} = m \times \left( T_d + T_n + \frac{T_c + T_s}{2} \right) + (n - m) \times T_B \quad (11)$$

### 3.1.3 Redundancy

The last factor redundancy,  $R$ , is the amount of spare resources reserved for recovery purpose. In the basic architecture with non-protection, the redundancy:  $R_B = 0$ . The redundancy of the Random selected protection scheme:  $R_R = \sum_{i \in M} (C_{Fiber_i} + C_{ONU_i})$ ,

where  $C$  is the spare capacity in link or component. The redundancy of Weakest and Strongest protection scheme which is represented as  $R_W$  and  $R_S$ , respectively, is similar to  $R_R$ , but the  $i \in M$  is selected in different way. Due to the Optical and wireless protection scheme adding additional fiber, the formula of its redundancy is:

$$R_{O\&W} = \sum_{i \in M} (C_{Fiber_i} + C_{ONU_i}) + C_{ONU_1} + C_{Fiber_1} + C_{Fiber_{back-up}} \quad (12)$$

## 3.2 Normalization and Application

To compare the protection schemes, after the values of each factor are calculated, normalization is performed. This process deals with the values in the mathematical sense, and places them in the [0,1] interval. As a result, normalized parameters are acquired:  $Q_A$ ,  $Q_T$ ,  $Q_R$ . Here, we use the shape functions [4].

$$Q_A = \frac{1}{1 - A_t^q + (1 + A_t - A)^q} \quad (13)$$

where  $q$  is shape parameter of the function, and  $A_t$  is the threshold value.  $A$  is the average value of each protection scheme equals:  $A = \frac{1}{n} A_i, i \in \{B, R, S, W, O \& W\}$ .

$$Q_T = \frac{1}{1+r \times T} \quad (14)$$

where  $r$  is the scale parameter, and  $T = \frac{1}{n} T_i, i \in \{B, R, S, W, O \& W\}$ .

$$Q_R = \frac{1}{1 + \frac{1 - Q_{R_i}}{Q_{R_i} \times R_{N_i}} \times R^3} \quad (15)$$

where  $Q_{R_i}$  is the value at  $R=R_i$ . Similarly,  $R = \frac{1}{n} R_i, i \in \{B, R, S, W, O \& W\}$ .

$$QoR = \frac{\alpha_A Q_A + \alpha_T Q_T + \alpha_R Q_R}{\alpha_A + \alpha_T + \alpha_R} \quad (16)$$

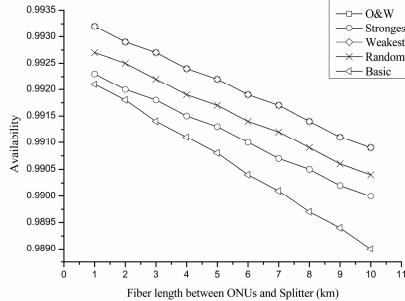
where  $\alpha_A, \alpha_T, \alpha_R$  are the weights of  $Q_A, Q_T, Q_R$ , respectively. The weights can be set arbitrarily in the different situations.

## 4 Numerical Results

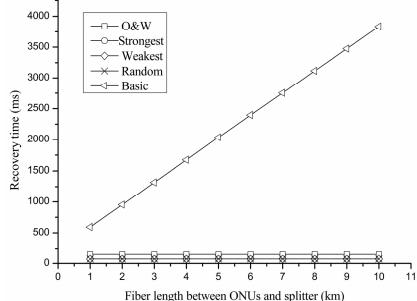
In the simulated WOBAN, the total ONUs number was  $n=32$  and the number of selected ONUs was  $m=8$ . Due to the different protection focusing on the selections among the different ONU<sub>i</sub>, the values of availabilities of ONU<sub>i</sub> was supposed as a descending order, in other words,  $A_{ONUi+1} > A_{ONUi}$ .

The availabilities ( $Q_A$ ) for different schemes related to the corresponding fiber length were shown in Fig.6. As it could be predicted, the basic scheme had the worst performance and it decreased rapidly with the length of fiber. The Weakest selection scheme overlaps with Optical and wireless scheme that both had the best performance. Among the protection schemes, the Strongest selection did not get such effective when compared with others. Because the strong ONUs had a high availability and combining them together was hard to improve the whole structure greatly. On the contrary, the combination of the weak ONUs can raise the availabilities of these ONUs from a low values so that the availability of overall ONUs improved.

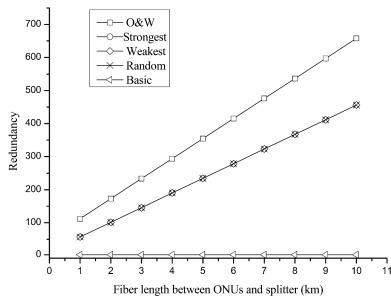
Next, we investigated the recovery time ( $Q_T$ ) of different protection schemes in Fig.7. We can see that the architecture with no protections had a long recovery time. Another hand, the recovery time with protections is much shorter. In the figure, the schemes of Random, Weakest and Strongest selection overlap with each other, because they were all wireless protection scheme just differ by selections of corresponding ONUs. From the figure, we knew that choosing the Optical and wireless scheme did not take much more time than the wireless ones. And the extra high time was cost by the switching the backup fiber.



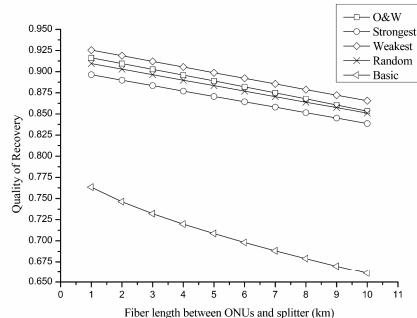
**Fig. 6.** The comparison of the availabilities of different schemes



**Fig. 7.** The comparison of the recovery times of different schemes



**Fig. 8.** The comparison of redundancies of different schemes



**Fig. 9.** The comparison of QoR calculated from different schemes

Fig. 8 depicted the redundancies for different schemes. For no extra resource reserved in the basic architecture,  $Q_R=0$ . The result showed that much more resources were used in the protection scheme as the distance increased. The wireless protections overlapped with each other like the front paragraph says. And the Optical and wireless scheme need an extra resource to the wireless scheme.

The results of QoRs were presented in Fig. 9. QoR for basic scheme was worst, and fall down quickly with the distance increasing. On the contrary, the QoRs of architecture with protection were much higher than it. Among the different wireless selection schemes, the Weakest selection scheme had the best performance with low redundancy. The Optical and wireless protection was a bit lower than the Weakest one, because of the extra backup fiber bringing a extra cost. And as the fiber length increased, the cost of the backup fiber caused a reducing of the QoR.

## 5 Conclusion

In this paper, four different ONU protection schemes based on a typical WOBAN structure have been compared. We emphasize assessing Quality of Recovery (QoR)

of the different schemes on availability, recovery time and redundancy. On the basis of the results, the weakest selection protection scheme is considered as the scheme with best cost effective.

**Acknowledgement.** This work was supported by the Natural Science Foundation of China (No. 61172081), the Natural Science Foundation of Zhejiang Province (No. LQ12F05008) and Natural Science Foundation of Zhejiang University of Technology (No.2011XY027).

## References

1. Feng, T.M., Ruan, L.: Design of a Survivable Hybrid Wireless-Optical Broadband-Access Network. *J. of Opt. Comm. and Netw.* 3, 458–464 (2011)
2. Navid, G., Michael, S., Martin, M.: Survivability Analysis of Next-Generation Passive Optical Networks and Fiber-Wireless Access Networks. *IEEE Trans. Reliab.* 60, 479–492 (2011)
3. Correia, N., Coimbra, J., Schutz, G.: Fault-Tolerance Planning in Multiradio Hybrid Wireless-Optical Broadband Access Networks. *J. of Opt. Comm. and Netw.* 1, 645–654 (2009)
4. Choda, P., Jajszczyk, A., Wajda, K.: A unified quality of recovery (QoR) measure. *Int. J. Commun. Syst.* 21, 525–548 (2008)
5. Choda, P., Wajda, K., Jajszczyk, A., Tapolcai, J., Cinkler, T., Bodamer, S., Colle, D., Ferraris, G.: Considerations about service differentiation using a combined QoS/QoR approach. In: Proc. DRCN Int. Workshop Design Reliable Com. Netw. Reliable Serv. 2005, pp. 345–352 (2005)
6. Sebbah, S., Jaumard, B.: Differentiated Quality-of-Recovery in Survivable Optical Mesh Networks Using p-Structures. *IEEE ACM Trans. Networking* (2011)
7. Choda, P., Jajszczyk, A., Kantor, M.: Quality of recovery (QoR) of access networks based on PON. In: Conf. Next Generation Internet Des. Eng., pp. 184–191 (2006)
8. Tapolcai, J., Choldat, P., Cinkler, T., Wajdat, K., Jajszczyk, A., Autenrieth, A., Bodamer, S., Colle, D., Ferraris, G., Lnsethagen, H., Svinnset, I., Verchere, D.: Quality of Resilience (QoR): NOBEL approach to the multi-service resilience characterization. In: 2nd Int. Conf. Broadband Networks BROADNETS 2005, pp. 405–414 (2005)
9. Cholda, P., Jajszczyk, A.: Recovery and Its Quality in Multilayer Networks. *J. Lightwave Technol.* 28, 372–389 (2010)

# Energy Efficient Hop Length Optimization for Laterally Connected Wireless Sensor Networks

Yuyang Peng, Ushan He, and Jaeho Choi

Dept. of Electronic Engineering, CAIIT, Chonbuk National University  
Chonju, Chonbuk, Republic of Korea 561-756  
*{Yuyang, wave}@jbnu.ac.kr, colinmengyu@hotmail.com*

**Abstract.** It is well known that energy efficiency is the main concern in wireless sensor networks due to the fact that sensors are frequently too small to carry enough energy for long-term operation. Here, an energy efficient hop length optimization scheme for the sensors in a wireless sensor network is presented. Particularly, the focus of the investigation is on the lateral connection architecture between the sensors of the network. For the verification of the proposed scheme, the total energy consumption for per bit is derived and compared to the equidistance scheme. The results obtained from numerical analyses indicate that the proposed scheme significant saves total energy consumption and can elongate the lifetime of the wireless sensor network.

**Keywords:** Wireless sensor network, Lateral connection sensor network architecture, Hop length optimization, Energy consumption.

## 1 Introduction

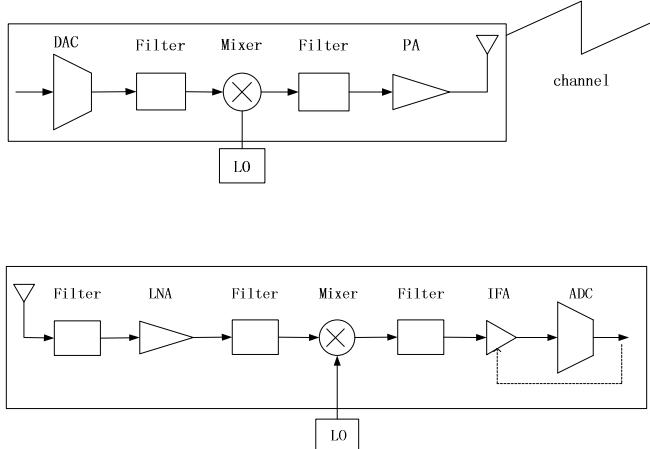
Nowadays, multi-hop transmission is gaining much attention in the communications and network research societies. Among them, an outstanding problem is the energy consumption of a wireless sensor network. The energy consumption of a wireless sensor network can be affected by many factors. They include network topology management, medium access control method, and error control schemes. In recent years, much research effort is made to design on optimal sensor placement schemes.

The wireless sensors can generally be placed in an area of interest either deterministically or randomly. The network connection architecture depends on the type of sensors, applications, and environments that the sensors will operate in. Selecting the hop-length between sensor nodes is viable and necessary when the network operation is not affected by the position of sensor nodes. Such scenarios include forest fire detection application in which an arbitrary deployment is available and the main concern is the life time of the whole sensor network. This paper focuses on prolonging the sensor network lifetime by optimizing the hop-length. In this investigation, the sensors are connected laterally and they deliver not only their own messages but the messages from the nodes in the downlinks to the sink of the wireless sensor network.

The paper is organized as follows. In Section 2 the system model of a sensor is introduced and the energy consumption per bit is discussed. In Section 3, the laterally connected wireless sensor network is considered and the optimization of the hop length between the sensors is investigated. The performance is verified by using numerical analysis and the results are compared to the conventional, equidistance scheme in Section 4; and, finally, the conclusion is made in Section 5.

## 2 System Model for Sensors and Energy Consumption

From [1] there are two main components accounting for the total average power consumption of a sensor. They are the power consumption in the power amplifier  $P_{PA}$  and the power consumption in the circuit blocks  $P_c$ . The transmitter and receiver circuit blocks of a sensor are shown in Fig. 1.



**Fig. 1.** The transmitter and receiver circuit blocks of a sensor

The  $P_{PA}$  is depend on the transmit power  $P_{out}$ , which can be calculated according to the link budget relationship [2]. Assuming the channel undergoes a square-law path loss, the transmit power can be obtained as follows:

$$P_{out} = \bar{E}_b R_b \times \frac{(4\pi d)^2}{G_t G_r \lambda^2} M_l N_f \quad (1)$$

where  $\bar{E}_b$  is the required energy per bit at the receiver for a given bit error rate (BER) requirement;  $R_b$  is the bit rate;  $d$  is the transmission distance;  $G_t$  and  $G_r$  are the transmitter and receiver antenna gains, respectively;  $\lambda$  is the carrier wavelength;  $M_l$  is

the link margin compensating the hardware process variations; and  $N_f$  is the receiver noise. It should be noted that  $N_f$  is given by  $N_f = N_r/N_0$  where  $N_r$  is the power spectral density (PSD) of the total effective noise at the receiver input and  $N_0$  is the single-sided thermal noise PSD at a room temperature with a value  $N_0 = -171$  dBm/Hz.

The power consumption of the power amplifiers can be approximated as follows [3]:

$$P_{PA} = (1 + \alpha) P_{out} \quad (2)$$

where  $\alpha = \xi/\eta - 1$  with  $\xi$  is the peak-to-average ratio (PAR) and  $\eta$  is the drain efficiency of the RF power amplifiers [4]. The total energy consumption per bit for a system can be obtained as follows:

$$E_{bt} = (P_{PA} + P_c) / R_b \quad (3)$$

where  $R_b$  is the data rate and from Eqs.(1), (2), and (3), we can get the energy consumption per bit as follows:

$$\bar{E}_{bt} = (1 + \alpha) \bar{E}_b \times \frac{(4\pi d)^2}{G_t G_r \lambda^2} M_t N_f + P_c / R_b \quad (4)$$

In Eq. (4),  $\bar{E}_b$  is defined by the BER. By adopting BPSK modulation in the Alamouti scheme, the average BER can be defined as follows [1]:

$$\bar{P}_b \approx \mathbb{E}_H \left[ Q(\sqrt{2\gamma_b}) \right] \quad (5)$$

$\mathbb{E}_H [ ]$  denotes the expectation considering the channel  $\mathbf{H}$ ,  $\gamma_b$  is the instantaneous received signal-to-noise ratio (SNR),  $Q( )$  is the  $Q$ -function. The average BER is upper-bounded by the Chernoff bound as follows [1]:

$$\bar{P}_b \leq \left( \frac{\bar{E}_b}{M_t N_0} \right)^{-M_t} \quad (6)$$

where  $M_t$  is the number of transmitter antenna and the required energy per bit is upper-bounded as follows:

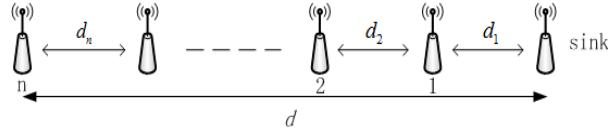
$$\bar{E}_b \leq \frac{M_t N_0}{\bar{P}_b^{1/M_t}} \quad (7)$$

By substituting Eq. (7) into Eq. (4), we can determine the total energy consumption per bit as follows:

$$\bar{E}_{bt} = (1 + \alpha) \frac{M_t N_0}{\bar{P}_b^{1/M_t}} \times \frac{(4\pi d)^2}{G_t G_r \lambda^2} M_t N_f + P_c / R_b \quad (8)$$

### 3 Laterally Connected Sensor Network and Hop Lengths

Now, consider a wireless sensor network with  $N$  sensors. Here, the sensors are laterally connected each other to deliver the messages of their own and of others in the downlinks to the sink of the network as shown in Fig. 2.



**Fig. 2.** Lateral connection wireless sensor network architecture

The hop length between sensor nodes is defined as  $d_i$  ( $i=1, 2\dots n$ ). The distance between the sensor nodes is the major consideration. For a transmission distance  $d_i$ , the energy consumption per bit can be defined as follows:

$$E_{bt}(d_i) = (1+\alpha) \frac{M_t N_0}{P_b} \times \frac{(4\pi d_i)^2}{G_t G_r \lambda^2} M_t N_f + P_c / R_b \quad (9)$$

Assuming each sensor node has an equal amount of data to transmit, the total energy consumption of the whole sensor network can be expressed as follows:

$$E_{total} = \sum_{i=1}^n (n+1-i) E_{bt}(d_i) R_b \quad (10)$$

In order to minimize the total energy consumption in terms of transmission distance, the distance  $d_i$  needs to be optimized. Constraining  $\sum_{i=1}^n d_i = d$ , the total energy consumption taking account of the distance can be defined as follows:

$$\Phi = \sum_{i=1}^n (n+1-i) E_{bt}(d_i) R_b + w(d - \sum_{i=1}^n d_i) \quad (11)$$

By taking partial derivatives on Eq. (11) with respect to  $d_i$ , a set equations is obtained as follows:

$$\frac{\partial \Phi}{\partial d_i} = 2E(n+1-i)d_i - w = 0 \quad (12)$$

where  $E = (1+\alpha) \frac{M_t N_0}{P_b} \times \frac{(4\pi)^2}{G_t G_r \lambda^2} M_t N_f R_b$  and  $w$  is a Langrange's multiplier.

Solving Eq.(12) each distance  $d_i$  can be determined as follows;

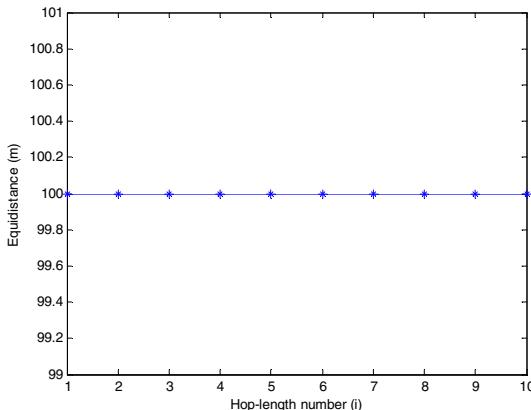
$$d_i = \frac{w}{2E(n+1-i)} \quad (13)$$

Since  $\sum_{i=1}^n d_i = d$ , Eq. (13) can be rewritten as follows:

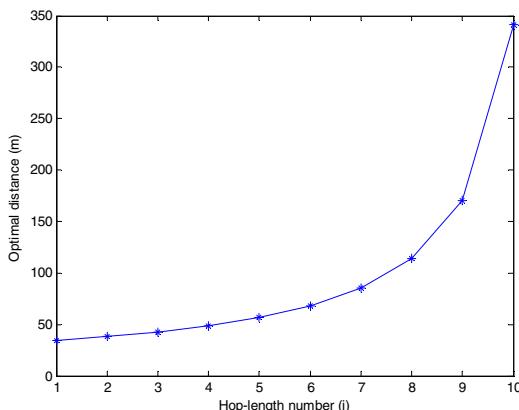
$$d_i = \frac{d}{\sum_{f=1}^n \left(\frac{1}{n+1-f}\right)(n+1-i)} \quad (14)$$

## 4 Results Using Numerical Analysis

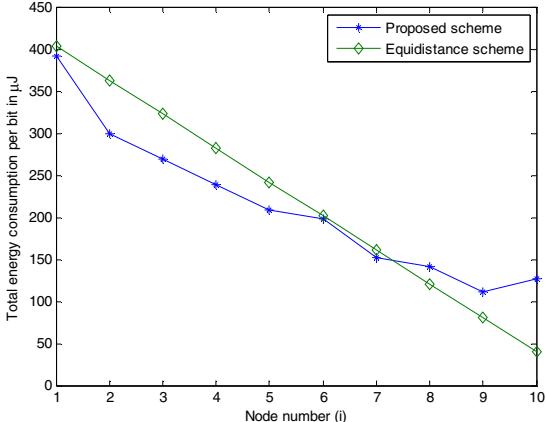
There are various scenarios for numerical analyses. Here, a typical example with  $n=10$ ,  $d=1000m$  is used to show the performance of the proposed scheme. Fig. 3 shows the conventional, equidistance scheme.



**Fig. 3.** Equidistance vs. hop-length number



**Fig. 4.** Optimal distance vs. hop-length number



**Fig. 5.** Total energy consumption per bit for each scheme ( $n=10$ ,  $d=1000\text{m}$ ,  $d_{long}=2\text{m}$ )

Fig. 4 shows the optimal hop distances between the wireless sensors in the laterally connected network architecture. As getting nearer to the destination sink, the hop-lengths between the sensors get smaller and smaller.

The reduction in total energy consumption can be achieved using the proposed scheme over the equidistance scheme. In this simulation, the following typical values are used.  $B=10 \text{ kHz}$ ,  $f_c=2.5 \text{ GHz}$ ,  $P_{mix}=30.3 \text{ mW}$ ,  $P_{fil}=2.5 \text{ mW}$ ,  $P_{fh}=2.5 \text{ mW}$ ,  $P_{LNA}=20 \text{ mW}$ ,  $P_{synth}=50 \text{ mW}$ ,  $M_f=40 \text{ dB}$ ,  $N_f=10 \text{ dB}$ ,  $G_rG_t=5 \text{ dBi}$ , and  $\eta=0.35$ ,  $N_i=20 \text{ kb}$ , and  $n_r=10$  [5]. In Fig. 5 the total energy consumption per bit is plotted for the proposed scheme as well as the conventional equidistance scheme. The results show that the proposed, optimal-distance scheme can save the energy consumption of the wireless sensor network. However, there is also the performance turning point in the multi-hop connection architecture. The energy-saving capability degrades gracefully when the multi-hopping number is more or less than seven as also shown in Fig. 5. Overall, the results show that the proposed scheme offers a total energy saving of about 4.5%.

In the lateral connection architecture considered here, regardless of two distance schemes compared, recall that the nodes closer to the sink have to carry heavier traffic loads than the ones located farther away; and the nodes that are closer to the sink will die first. Concerning the lifetime, the proposed method can elongate the network lifetime about 3% longer than the equidistance scheme using the same parameters above.

## 5 Conclusion

A hop length optimization scheme has been proposed. The proposed hop distance scheme is capable of minimizing the energy consumption of sensors in wireless sensor network where the sensors are making laterally connected

multi-hop network architecture. The results demonstrate that the proposed scheme offers a total energy saving of around 4.5% and a network lifetime saving of 3% in comparison to the conventional equidistance scheme. The future research is considered to investigate the hop distance problem the network topology changes into a two-dimensional form such as a mesh.

## References

1. Cui, S., Goldsmith, A., Bahai, A.: Energy-efficiency of MIMO and cooperative MIMO techniques in sensor network. *IEEE J. Select. Areas Comm.* 22(6), 1089–1098 (2004)
2. Proakis, J.: *Digital Communications*, 4th edn. McGraw-Hill, New York (2000)
3. Cui, S., Goldsmith, A., Bahai, A.: Modulation optimization under energy constraints. In: Proc. of ICC 2003, pp. 2805–2811 (2003)
4. Lee, T.: *The Design of CMOS Radio-Frequency Integrated Circuits*. Cambridge Univ. Press, Cambridge (1998)
5. Cui, S., Goldsmith, A., Bahai, A.: Energy-constrained modulation optimization. *IEEE Trans. Wireless Comm.* 4, 2349–2360 (2005)

# Performance Evaluation Analysis about Ethernet and DeviceNet

Wen Li<sup>1</sup> and Xiangyu Dai<sup>2</sup>

<sup>1</sup> Hunan Institute of Science and Technology, Yueyang, Hunan, 414006, China  
liwen07@sohu.com

<sup>2</sup> Changsha Professional Institute of Electric and Technology,  
Changsha, 410131, Hunan, China  
daixiangyu09@gmail.com

**Abstract.** The mechanisms of medium access control (MAC) layer about Ethernet and DeviceNet are discussed and the Ethernet bus with carrier sense multiple access with collision detection(CSMA/CD) and DeviceNet with token-passing bus are compared in detail in this paper. The emulated result shows the transmission time of Ethernet is lower than DeviceNet but the transmission time of Ethernet almost is constant in little data capacity and the coding efficiency of Ethernet is lower than DeviceNet in little data capacity but in big data capacity the coding efficiency of Ethernet is higher. Its transmission time and data coding efficiency are emulated in MATLAB.

**Keywords:** CSMA/CD, Evaluation, Transmission Time, Data Coding Efficiency.

## 1 Introduction

In the present time, although many field bus, such as FF, CAN, Lonworks, PROFIBUS, can provide good real time characteristic, they can not be unified in the near future due to their own characteristics and applying areas, along with many economic and political reasons. Meanwhile Ethernet becomes the popular network architecture because of its low cost and efficiency, and it is widely used in computer network. Theoretically, Ethernet seems should be the direction of the field bus. But Ethernet is not a real time network. It is urgent to resolve the real time issue which needs to start with the analysis of MAC layer protocol. In this article, we will discuss Ethernet bus with carrier sense multiple access with collision detection (CSMA/CD) and ControlNet with token-passing bus. The comparative analysis of the mechanisms of network control, the medium access control layer protocol of both Ethernet and ControlNet will be explored, as well as the data transmission time and the data coding efficiency.

## 2 The Mechanism of MAC Layer of Ethernet and ControlNet

In this section, we will discuss the MAC layer protocol of two control networks – Ethernet and ControlNet. Ethernet is not a complete protocol, it's just a definition of a

MAC layer, but ControlNet is. The MAC Layer manages and maintains communications between access nodes by coordinating access to a channel and utilizing protocols that enhance communications over a wireless medium. The MAC Layer coordinates real time control between multiple access nodes, and it is responsible for the communications quality and reliability. The comparative analysis of the mechanism of the MAC layers of these control networks will benefit the research toward the enhancement of the real time functionality of the Ethernet network.

Ethernet uses the Carrier Sense Multiple Access with Collision Detection (CSMA/CD) mechanism for resolving contention on the communication medium, which is specified in IEEE 802.3 network standard. When a node wants to transmit any data, it listens to the network. If the network is busy, it refrains from transmitting until the network quiets down; otherwise it transmits immediately. If two or more nodes listen to the idle network and decide to transmit simultaneously, the messages of these transmitting nodes collide and the messages are corrupted. With this strategy, the nodes continue to listen to the network as they transmit. If a node detects another signal that interferes with the signal it's sending, it stops transmitting. Both nodes then wait a random amount of time and attempt to retransmit. This random time is determined by the standard Binary Exponential Backoff (BEB) algorithm. The time before trying again is randomly chosen between 0 and  $(2i-1)$  slot times, where  $i$  denotes the  $i$ th collision event detected by the node and one slot time is the minimum time needed for a round-trip transmission. However, after 10 collisions have been reached, the interval is fixed at a maximum of 1023 slots. After 16 collisions, the node stops attempting to transmit and reports failure back to the node microprocessor. Further recovery may be attempted in higher layers.

Because of low medium access overhead, Ethernet uses a simple algorithm for operation of the network and has almost no delay at low network loads. No communication bandwidth is used to gain access to the network compared with the token bus or token ring protocol. Ethernet used as a control network commonly uses the 10 Mbps standard; high-speed (100 Mbps or even 1 Gbps) Ethernet is mainly used in data networks.

Ethernet is a nondeterministic protocol and does not support any message prioritization. At high network loads, message collisions are a major problem because they greatly affect data throughput and time delay which may be unbounded. The Ethernet capture effect existing in the standard BEB algorithm, in which a node transmits packets exclusively for a prolonged time despite other nodes waiting for medium access, causes unfairness, and results in substantial performance degradation. Based on the BEB algorithm, a message may be discarded after a series of collisions; therefore, end-to-end communication is not guaranteed. Because of the required minimum valid frame size, Ethernet uses a large message size to transmit a small amount of data.

Several solutions have been proposed for using Ethernet in control applications. For example, every message could be time-stamped before it is sent. This requires clock synchronization, however, which is not easy to accomplish, especially with a network of this type. Various schemes based on deterministic retransmission delays for the collided packets of a CSMA/CD protocol result in an upper-bounded delay for all the

transmitted packets. However, this is achieved at the expense of inferior performance to CSMA/CD at low to moderate channel utilization in terms of delay throughput. Other solutions also try to prioritize CSMA/CD (e.g., LonWorks) to improve the response time of critical packets. Using switched Ethernet by subdividing the network architecture is another way to increase its efficiency.

MAP, PROFIBUS, and ControlNet are typical examples of token-passing bus control networks. These are deterministic networks because the maximum waiting time before sending a message frame can be characterized by the token rotation time. The token bus protocol (IEEE 802.4) allows a linear, multi-drop, tree-shaped, or segmented topology. The nodes in the token bus network are arranged logically into a ring, and, in the case of ControlNet, each node knows the address of its predecessor and its successor. During operation of the network, the node with the token transmits data frames until either it runs out of data frames to transmit or the time it has held the token reaches the limit. The node then regenerates the token and transmits it to its logical successor on the network. If a node has no message to send, it just passes the token to the successor node. The physical location of the successor is not important because the token is sent to the logical neighbor. Collision of data frames does not occur, as only one node can transmit at a time. The protocol also guarantees a maximum time between network accesses for each node, and the protocol has provisions to 5 regenerate the token if the token holder stops transmitting and does not pass the token to its successor. Nodes can also be added dynamically to the bus and can request to be dropped from the logical ring.

The ControlNet protocol adopts an implicit token-passing mechanism and assigns a unique MAC ID (from 1 to 99) to each node. As in general token-passing buses, the node with the token can send data; however, there is no real token passing around the network. Instead, each node monitors the source MAC ID of each message frame received. At the end of a message frame, each node sets an implicit token register" to the received source MAC ID + 1. If the implicit token register is equal to the node's own MAC ID, that node may now transmit messages. All nodes have the same value in their implicit token registers, preventing collisions on the medium. If a node has no data to send, it just sends a message with an empty Lpacket field, called a null frame.

The length of a cycle, called the Network Update Time (NUT) in ControlNet or Token Rotation Time (TRT) in general, is divided into three major parts: scheduled, unscheduled, and guardband. During the scheduled part of a NUT, each node can transmit time-critical/scheduled data by obtaining the implicit token from 0 to S. During the unscheduled part of a NUT, each node from 0 to U shares the opportunity to transmit non-time-critical data in a round-robin fashion until the allocated unscheduled duration is expired. When the guardband time is reached, all nodes stop transmitting, and only the node with lowest MAC ID, called the "moderator," can transmit a maintenance message, called the "moderator frame," which accomplishes the synchronization of all timers inside each node and publishing of critical link parameters such as NUT, node time, S, U, etc. If the moderator frame is not heard for two consecutive NUTs, the node with the lowest MAC ID will begin transmitting the moderator frame in the guardband of the third NUT. Moreover, if a moderator node

notices that another node has a lower MAC ID than its own; it immediately cancels its moderator role.

The token bus protocol is a deterministic protocol that provides excellent throughput and efficiency at high network loads. During network operation, the token bus can dynamically add nodes to or remove nodes from the network. Scheduled and unscheduled segments in each NUT cycle make ControlNet suitable for both time-critical and non-time-critical messages. Although the token bus protocol is efficient and deterministic at high network loads, at low channel traffic its performance cannot match that of contention protocols. In general, when there are many nodes in one logical ring, a large percentage of the network time is used in passing the token between nodes when data traffic is light.

### 3 Comparative Analysis of Two Control Networks

The nodes in the Ethernet network use CSMA/CD mechanism for resolving contention on the communication medium. As the frame arriving process is compiled with Poisson's Distribution the arrival interval should be complied with  $1/\lambda$  power distribution. When a node detects a busy network or a data collision, the delay interval of retransmit is compiled with  $1/v$  power distribution. In the simulation process, random numbers must be generated first. Assume that the frame arrival interval is set as  $t$ , and then the distribution formula for arrival interval can be express as

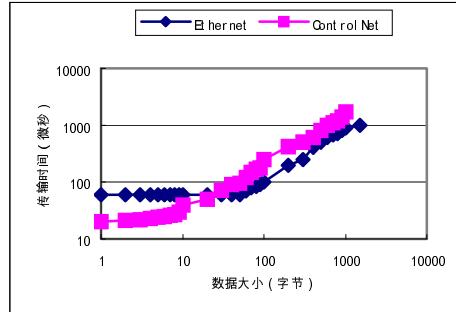
$$P(\tau) = \begin{cases} 1 - e^{-\lambda\tau} & (\tau \geq 0) \\ 0 & (\tau < 0) \end{cases} \quad (1)$$

For the random numbers  $X$  ( $0 < X < 1$ ), which is compiled with Even Distribution,  $X = 1 - P(\tau)$ ,  $\tau = -\ln X / \lambda$ . In other work, as long as we can generate radon numbers compiled with Even Distribution, we can produce radon numbers compiled with Power Distribution. By using Recursion Rule, then false random numbers compiled with Even Distribution between 0 and 1 can be generated.

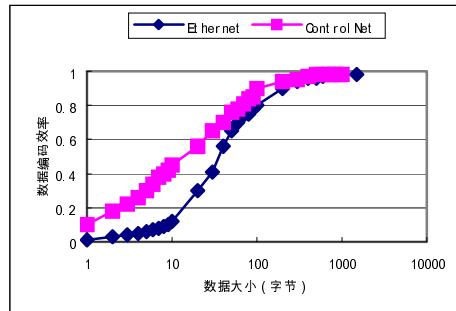
In the simulation process, set  $X_i = 18687X_{i-1} - \text{TRUNC}(18687X_{i-1})$ , where  $i = 1, 2, \dots$ ,  $X_0 = 1/29657$ . Random numbers complied with Power Distribution and with medium  $= 1/\lambda$  can be generated from  $X_i$ . From the same principal, random numbers complied with Power Distribution and with medium  $= 1/v$  can also be generated. Token passing network is deterministic, so the message frame is governed by its protocol.

We used MATLAB simulator to compare these two control networks, in the aspects of data transmission time and data coding efficiency. In the simulation, For Ethernet, the data transmission rate is 10 Mbps, the maximum distance is 2500 meter and the maximum data length is 1500 byte; For ControlNet, the data transmission rate is 5 Mbps, the maximum distance is 1000 meter and the maximum data length is 504 byte. The average node number is 8. The transmission time is determined with the average time consumed to transmit a certain bytes of data successfully. Data coding efficiency is defined by the ratio of data size and the message size. The message size is total number of bytes in the data transmission.

As shown in Fig. 1, the transmission time for ControlNet is less than Ethernet on small data size. Ethernet requires less transmission time on larger data sizes ( $>50$  bytes). Although ControlNet uses less time to transmit the same size of small data, it needs some amount of time (NUT) to gain access to the network. At the same time, we observed that non-real time Ethernet has very good determination of transmission time what the data size is small.



**Fig. 1.** Comparison of the data size vs. the coding efficiency



**Fig. 2.** Comparison of the data size vs. data transmission time

Fig. 2 shows the relationship between the data size and coding efficiency. When the data size is small, the ControlNet is better; when the data size is large, they are 98% similar. For the ControlNet, the data size is normally smaller.

The Inflection points in the Fig. 1 and Fig. 2 are caused by data section, for example, the maximum data size. The smooth area in Fig. 1 is caused by the small size of data.

## 4 Conclusion

Through the above analysis, we proved that the characteristics of Ethernet are very encouraging for the smaller data control network. Especially the transmit time is constant for the small data size. It is perspective to enhance the real time characteristics

of Ethernet, so that it can fit for industrial process control and unify the bus standard. Also it set as a solid foundation for further research on network average delay time, total delay time, and network throughput. Although Ethernet frame arrival rule can be proved strictly, the approximate in the modeling process can cause the difference between the simulation result and actual network transmission process. This difference can be eliminated in the future experiment for the improved Ethernet protocol.

**Acknowledgment.** This project supported by Hunan Provincial Natural Science Foundation of China (10JJ2044).

## References

1. Zengwenbo: Control network technology. Tsinghua university Press. Springer, Beijing (2011)
2. Spurgeon, C.E.: Ethernet technology. Tsinghua university press, Beijing (1998)
3. Gongzhong: MATLAB and emulate. Tsinghua university press, Beijing (2010)
4. Lian, F.-L., Moyne, J.: Control performance study of a networked machining cell. In: American Control Conf., pp. 2337–2341 (June 2000)

# Network Interaction Analysis of 3G and WiFi

Xuejun Meng

Network Center,Wuhan University  
Wuhan, China  
xjmeng@whu.edu.cn

**Abstract.** The sufficient coalescent of high speed rate and agility for WLAN with high dependability and roam function for 3G make network interaction possible to satisfy people's need. The paper carried on a comparison of technique characteristics of WiFi and 3G and particularly discussed various network mutual communication methods of two kinds of techniques, analyzed EAP-AKA safety authentication mechanism.

**Keywords:** 3G, WiFi, EAP-AKA, Authentication.

## 1 Introduction

3G is a standard of wireless wide area network (WWAN), mainly based on wide area overlay of base station. Its characteristics is wide overlay scope of network, provision of businesses such as voice, data and multimedia etc., strong roam function, and high security, but its disadvantage is lower baud rate. According to WLAN standard, WiFi mainly used for distributed connection of close distance about Internet/Intranet in hot spot region, can carry on switch at lower speed under control of AP's controller, is a kind of mobile, flexible and fast system. Therefore, connection of WLAN and 3G with each other will make two kinds of technical advantages get full exertion.

## 2 Technique Characteristics of 3G and WIFI

WLAN technique includes 802.11a/b/g/n, 3G technique mainly includes WCDMA, CDMA2000 and TD-CDMA. The comparison of technical characteristics about WLAN and 3G relatively refers to table 1.

The frequency segment of WLAN is 2.4GHz and 5.8GHz. 2.4GHz among those is open frequency segment, it is divided into 13 channels and we generally use 3 channels CH1, CH6 and CH11 in order to avoid the mutual interference among channels. 5.8GHz have 5 usable channels in 20MHz bandwidths. If the throughput of single channel is 20Mbit/s(802.11g), then the throughput of it in busyness is theoretic is 70Gbit/s/AP, but its value is approximate 30~40Gbit/s in the actual environment, this is more than throughput of 3G base station(for example, the throughput of EV-DO Rev.A base station of S111 is 13~20Gbit/s in busyness , average throughput is

**Table 1.** The Comparison of technical characteristics about WLAN and 3G

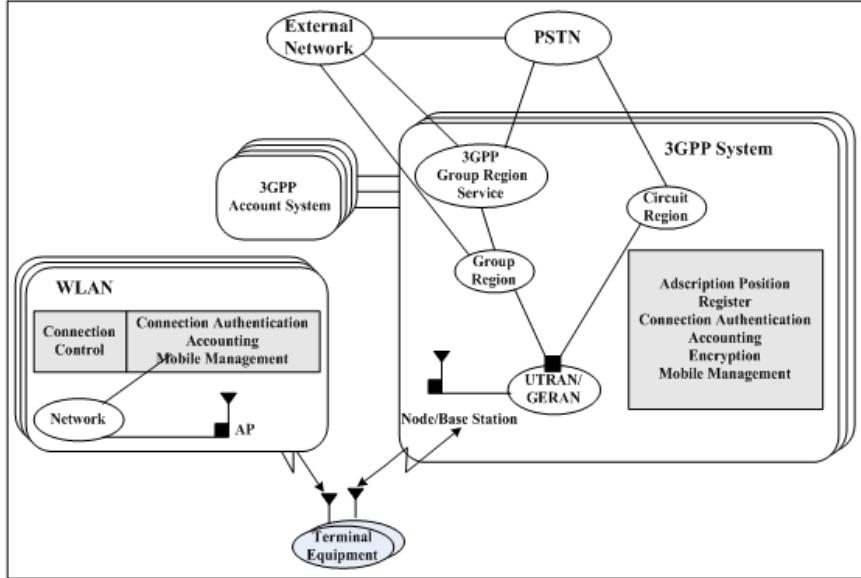
Technology	Criterion	Bandwidth	Type	Peak value of velocity	Range	Frequency
WiFi	802.11g	20MHz	LAN	22Mbit/s	300 feet	2.4GHz
	802.11n	20MHz	LAN	50Mbit/s		5.8 GHz
				100Mbit/s(2x2)		
3G	WCDMA	5MHz	WAN	2Mbit/s,10Mbit/s (technology of HSDPA)	1 ~ 5 miles	1.8GHz 1.9GHz 12.1GHz
	CDMA2000	1.23MHz	WAN	2.4Mbit/s, 300~600kbit/s		0.4/0.8/0.9/1.7/1.8 /1.9/2.1GHz
	TD-CDMA	1.6MHz	WAN	academic 2.8Mbit/s		155MHz

1.2~1.8 Mbit/s/carrier sector). The throughput of 802.11n, namely technique of Long Term Evolution is one time than that of 802.11g. Overlay scope of WLAN is smaller, mostly in 50~100 meters, generally not more than 50 meters. Due to adoption of MIMO 802.11n enhance its overlay ability, can reach to 100 meters. Therefore, due to high throughput, small overlay scope, WLAN is the best means that solves data business in hot spot according to the business capacity.

Because the design target of 3G is beehive communication technique that supports high speed ambulation, at design stage designer have already completely considered ambulation management(for example switch, Radio Paging, roam, register etc.), overlay ability, safety management, large-scale deployment ability and QOS etc., therefore the 3G technique obviously surpasses WLAN technique in these aspect. But about data throughput, in the center scope (<50 meters) that base station cover with, WLAN obviously surpasses 3G technique. Along with emergence of thin AP technical, AC (AP Controller,) can control AP, raise switch performance among APs and wireless resources management ability, improve QOS of WLAN.

### 3 Reference Model of 3G and WLAN Mutual Communication

Figure 1 show the reference model of mutual communication between 3G and WLAN. It mainly explained network unit that need to be introduced in mutual communication between 3G and WLAN. According to the dissimilarity of mutual communication method, it is not same either to the request of equipments. The function of the mutual communication that WLAN need includes connection control, connection authentication, accounting, ambulation management. The function of the mutual communication that 3G need includes to connection authentication, accounting, encryption, ambulation management.

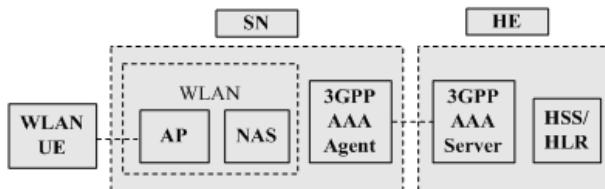


**Fig. 1.** The reference model of 3G and WLAN mutual communication

## 4 EAP-AKA Authentication Agreement Analysis

### 4.1 Interconnection Framework

3GPP (The 3rd Generation Partnership Project) aimed at 3G and WLAN network amalgamation to put forward a set of interconnection project and three kinds of interconnection framework. It depicts interconnection requirement, design EAP-AKA (Authentication and Key Agreement) according to safety connection of interconnection. The simplification diagram of 3G-WLAN interconnection framework is divided into three parts, such as figure 2 shows.



**Fig. 2.** Simplification diagram of interconnection framework of 3G-WLAN

Adscription Environment: HE mainly includes ownership user's server/ownership position register and 3GPP AAA server in 3G-WLAN Interconnection network.

Business Network: The function is responsible for overall scope of WLAN. It mainly includes 3GPP AAA proxy, Network Access server (NAS) and Access Point.

## 4.2 Safety Authentication Mechanism of EAP-AKA

EAP-AKA, namely Authentication and Key Agreement is based on USIM. USIM of customer and IILR of service provider share a key. The key is wrote a time in making USIM, is subjected to protection of safety mechanism of USIM. It can not be read. Consequently crack to USIM is very difficult. It is thought as a kind of authentication system of high safety strength in physical application.

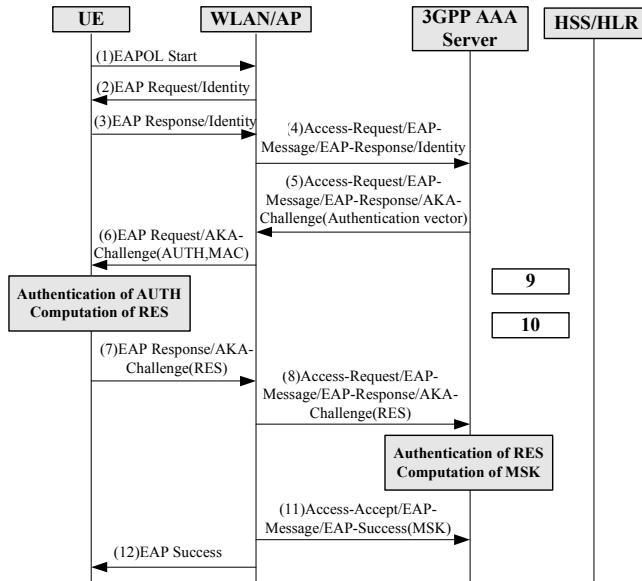
The EAP-AKA includes two kinds of authentication methods: full authentication and fast re-authentication. The full authentication divides two stages. AAA server with HSS/HLR executes AKA arithmetic to obtain AV in First stage, namely the AV obtainable stage. AAA server with WLAN terminal negotiates about double authentication and key material in second stage, namely AV usable stage. In the process of re-authentication key material of full authentication is used, there is no task about obtainment of AV. There is only AV usable stage, there is no AV obtainment stage.

In the 3GPP-WLAN AKA, each AV vector includes five parts: random number(RAND), authentication sign (AUTN), expectation response(XRES), integrity checkout key(IK), data encryption key(CK). Because authentication each time use AV vector, there is big expense for Signal transmission. In order to decrease the number of HSS and economize transmission spending that HSS obtains AV, the AAA server produce n AV vectors when it requests HSS (Home Subscriber Server)/HLR (Home Location Register) every time. But, to each authentication process, process should use AV vector to heavens so that it can reduce number that HSS/HLR produce AV and decrease correspondence spending.

Because of superiority of AKA authentication, the WLAN introduce it into arithmetic that verifies the customer identity. The WLAN usually adopt 802.1X and RADIUS system framework as network connection authentication, Authentication method in common use EAP-MD5, EAP-OTP, EAP-TLS etc.. Hence EAP-AKA authentication mechanism is produced. The AAA Server maintains all information of authentication and authority of customers, runs software of EAP-AKA server and is responsible for legitimacy verification of customers. Client is computer or phone supporting WLAN that equip UICC(the USIM integrated circuit card) and run client software of EAP-AKA.

Because EAP-AKA transfers messages according to the EAP agreement, two ends of authentication, client and server adopt EAP agreement for correspondence. EAP messages need to cross a few networks in the process of delivering from client to server. Firstly EAP messages that client send out deliver to the AP through the WLAN network. Secondly it delivers to the WLAN/AP through the ether net again. Finally it delivers to the AAA Server through an IP net, namely the EAP server end. The server that sends out EAP messages toward client experience contrary process. Therefore, WLAN, Ethernet and IP net between client and server need to provide loading for EAP message. The WLAN and Ethernet were similar to be two layer networks, and provide a same LLC (Logic Link Control) layer. Therefore, to the EAP agreement EAP message need to traverse IP net while it delivers from WLAN/AP to server. WLAN/AP and AAA Server need to adopt RADIUS agreement to deliver

authentication, authorization and accounting information. Therefore, RADIUS agreement can be used as loading, EAP message can be encapsulate in attribute of RADIUS message to deliver. EAP-AKA authentication process shows as figure 3. The concrete process is as follows:



**Fig. 3. EAP-AKA authentication process**

- (1) The client start authentication process by sending out EAPOL-Start message toward WLAN/AP.
- (2) WLAN/AP request to acquire customer's identifier by sending out EAP-Packet/EAP-Request message toward client.
- (3) The client reply customer's identifier (permanent identifier or temporary identifier).
- (4) WLAN/AP encapsulate EAP message in RADIUS message and send out Access-Request to the server.
- (5) According to authentication arithmetic, AAA server produce authentication vector, encapsulate authentication vector in EAP attribute, then encapsulate EAP message in RADIUS message and send out Access-Challenge message to WLAN/AP.
- (6) WLAN/AP withdraw EAP message from RADIUS message and deliver it to customer.
- (7) The customer validate legitimacy of network by using validation arithmetic, and compute out RES to send out to WLAN/AP.
- (8) WLAN/AP encapsulate EAP message in RADIUS message and send out to the server.
- (9) 3GPP AAA Server start checking whether the customer has N(2 or 3) available authentication 5 array(RAND, XRES, CK, IK, AUTN). If there is no enough 5 array, then send out MAP\_Send\_Auth\_Info message toward HLR by No.7 Signal to obtain N group authentication set. The purpose of using N 5 arrays is to produce longer Session Key.
- (10) 3GPP AAA Server still check whether the database has contract information about WLAN connection customer, if have no, then obtain from HSS/HLR, at the same time 3GPP AAA Server check if the customer has already

made a contract about WLAN business. (1) After server validate legitimacy of response, it send out Access-Accept message to WLAN/AP to notice authentication passing, and supplementary master conversation key(MSK). (2) WLAN/AP send out EAP-Success to client for notifying authentication passing. The client use key CK and IK to compute out MSK that is used as cipher key in safety system, such as 802.11 is and WPA etc..

## 5 Existent Problem

After many fulfillment and research, EAP-AKA authentication exists some loopholes and shortage as follows: (1) Authentication process need many times interaction of request and response with each other, this result in a little bit postponement of Authentication. (2) Because Internet Service Provider may own different wireless connection equipments, this will need additional trust management function. (3) Under the some condition, the system allows to pass IMSI in clear text to Authenticate user, this is safety loophole. (4) Network side don't need Authentication and connection point always be trusted. (5) EAP-AKA do not support encryption suite negotiation and Authenticate agreement edition negotiation. (6) EAP-AKA based on symmetry encryption system, it do not support not-symmetry encryption system and connection Authenticate based on certificate Authenticate.

## 6 Conclusion

Because of different technique characteristic of 3G and WLAN, both can help each other and can develop in phase and promote mutually. Especially after some problem about mutual communication, roam and Authentication between 3G and WLAN are solved, WLAN is used as bandwidth complementarity of hot spot district. This not only can consumedly save network construction cost, and also biggest enrich business application of future terminal. WLAN should have vaster development prospect.

## References

1. Wang, Z., Liu, W., Ke, Y.: Fast Authentication Method for 3G Users Among WLANs. Computer Engineering 36(10), 170–172 (2010)
2. Xu, W., Hou, H.: Analysis and Amendment of EAP-AKA Protocol in 3G-WLAN Interworking. Computer Engineering and Applications 46(32), 81–83 (2010)
3. Sun, H., Wang, Y., Xu, T., Zhang, L.: Improvement of EAP-AKA Protocol Signalling for 3G-WLAN. Computer Applications and Software 27(3), 13–15 (2010)
4. Wei, W., Li, Z.: Design Architectures for 3G and WLAN Integration. Production and Project 12, 57–59 (2005)
5. Wei, S.: 3GPP and WLAN Interworking Technology. Telcom Information (October 2004)

# Ant Intelligence Routing Algorithm for Wireless Sensor Networks

Awudu Karim<sup>1</sup>, Xiaoming Zhang<sup>2</sup>, A.M. Oluyemi<sup>3</sup>, and T. Fitarkandro<sup>3</sup>

College of Information Science and Engineering  
Hunan University, Lushan Road Changsha, 410082, China  
awudubody@yahoo.com

**Abstract.** In this paper, we present ant intelligence routing algorithm (AIRA), an adaptive, energy efficient and multiple-path protocol designed for wireless sensor networks. The primary goals of the protocol design are energy efficiency and self-organization without compromising throughput. AIRA reduces energy consumption by enabling low-duty-cycle operation and clocking neighbors to power of their radios to avoid unnecessary listening and interference during data transmission in a multihop network through adaptive sleeping technique. This greatly improves energy efficiency. It supports self-organization of individual nodes and reduces control overheads by using data packets themselves to maintain an established route for communication. Finally, AIRA applies synchronized sleeping technique to improve energy efficiency of the entire network. In an extensive set of simulations, we compare our routing algorithm with a state-of-the-art algorithm, and show that it gets better performance over a range of different scenarios.

**Keywords:** sensor networks, energy efficiency, ant colony based algorithms.

## 1 Introduction

Wireless sensor networking is an emerging technology that has a wide range of useful applications including monitoring the environment, security and warning systems, life-threatening places, medical systems, robotics, virtual environments, surveillance and so on. These networks may consist of large numbers of distributed sensing nodes that interrelate with themselves forming a multihop wireless network. A sensing node may consist of one or more sensors, embedded processors, memory and low-power radios. These nodes coordinate to perform a common task similar to ants in a colony working independently and cooperatively to achieve a common goal. Sensing nodes are typically battery powered and therefore are power limited. Due to their limited power, transmission range of wireless interfaces are also limited resulting in traffic being relayed over several intermediate nodes to enable communication between sender and receiver.

Modern communication networks are becoming increasingly diverse and heterogeneous. This is the consequence of the addition of an increasing array of devices

and services, especially wireless. The need for flawless interaction of numerous heterogeneous network nodes represents a formidable challenge.

The need to incorporate wireless networks into the existing wire-link infrastructure renders the requirement for efficient network routing even more demanding.

Current routing algorithms are not adequate in dealing with the increasing complexity of such networks. Centralized algorithms have scalability problems; static algorithms have trouble keeping up to date with network changes and on the other hand, distributed and dynamic algorithms have oscillations and stability problems.

Ant colony based routing provides a promising alternative to these approaches. It utilizes mobile software agents for network management. These agents are autonomous entities, both proactive and reactive, and have the capability to adapt, cooperate and move intelligently from one location to the other in the communication network. Ant colony exhibits emergent behavior whereby simple interactions of autonomous agents, with simple activities, give rise to a complex behavior that has not been specified explicitly.

There is currently an increasing interest for the paradigm of autonomic computing. This is basically because networks are becoming more and more complex and larger and that it is desirable that they can self organize and self-configure, adapting to new situations in terms of traffic, services, network connectivity, and so on. To support this new paradigm, network algorithms should be robust, work in a distributed way, be able to observe changes in the network and adapt to them without compromising connectivity.

Nature's self-organizing systems like ant colony show precisely these desirable properties. Making use of a number of relatively simple biological agents (e.g., the ants) a variety of different organized behaviors is generated at the system-level from the local interactions among the agents and with the environment. The robustness and effectiveness of such collective behaviors with respect to variations of environmental conditions are key aspects of their success. Nature's self-organizing systems have recently become a source of inspiration for the design of distributed and adaptive algorithms, and in particular routing algorithms.

## 2 Related Work

Several successful routing algorithms have been proposed taking inspiration from ant colony behavior and the related framework of Ant Colony Optimization (ACO) [1-16]. Examples of ACO routing algorithms are AntNet [2-9] and ABC [3-23]. The ACO routing algorithms mentioned were developed for wired networks. They work in a distributed and localized way, and are able to observe and adapt to changes in traffic patterns. However, changes in wireless networks are much more drastic. Topology and number of nodes can change continuously. Further difficulties are posed by the limited practical bandwidth of the shared wireless channel. Although the data rate of wireless communication can be quite high, algorithms used for medium access control, such as IEEE 802.11 DCF[4-14], create a lot of overhead both in terms of control packets and delay, lowering effectively the available bandwidth. The challenges of autonomic

networks are therefore much bigger, and new designs are necessary to guarantee even the basic network functions.

One type of networks where the need for autonomic control is essentially necessary is wireless sensor networks. These are networks in which all nodes communicate with each other via wireless connections. Nodes can join or leave at any time. All nodes are equal and there is no centralized control. There are no designated routers. Nodes serve as routers for each other, and data packets are forwarded from node to node in a multi-hop fashion.

This paper describes AIRA protocol designed for wireless sensor networks with the primary goals of energy efficiency and self organization while maintaining high throughput (end-to-end).

To achieve this, we need to identify the main sources that cause inefficient use of energy.

The first one is collision. When a transmitted packet collides with another, it has to be discarded and re-transmitted. This increases energy consumption as well as delay. The second source is overhearing, meaning that a node picks up packets that are destined to another node. The third source is control packet overhead. Sending and receiving control packets consume energy too. The last major source of energy inefficiency is idle listening, listening to possible traffic that is never sent. This is especially true in many sensor network applications. If nothing is sensed, nodes are in idle mode for most of the time. Researchers have shown that idle listening consumes 50%–100% of the energy required for receiving. For example, Stemm and Katz measured that the idle:receive:send ratios are 1:1.05:1.4 [5-15], while the Digitran wireless LAN module (IEEE 802.11/2 Mb/s) specification shows idle:receive:send ratios is 1:2:2.5 [6-24].

### 3 Background

Ant algorithms are multi-agent systems consisting of independent agents with individual ants behaviors. The ant colony optimization meta-heuristic is a particular class of ant algorithms.

The basic idea of the ant colony optimization Meta heuristic is taken from the food searching behavior of real ants. Ants on their way in search of food, start from their nest and move towards the food. When an ant reaches a junction, it has to decide which branch to take next. While moving, ants deposit pheromone, which marks the route taken. On their way back to the nest, the ants have to select a path with high trace of pheromone. After a short time the pheromone concentration on the shorter path will be higher than on the longer path, because the ants using the shorter path will increase the pheromone concentration faster. The shortest path will thus be identified and eventually all ants will only use this path. The concentration of pheromone on a certain path is an indication of its usage. Over time the concentration of pheromone decreases due to diffusion effects.

This behavior of the ants can be used to find the shortest path in networks. Especially, this method is dynamic and allows for a high adaptation to changes in

wireless networks, since in these networks the existence of links are not guaranteed and link changes occur rapidly.

The dynamic nature of wireless networks is responsible for the inadequate performance of several routing algorithms. The ant colony optimization meta-heuristic is based on agent systems and works with individual ants. This allows a high adaptation to the changes in the topology of a network.

Each node has a routing table with entries for all its neighbors, which contains also the pheromone concentrations. The decision to select the next hop is based on the pheromone concentration on the current node, which is provided for each possible link. Thus, the approach supports multipath routing.

In contrast to other routing approaches, the ant colony optimization meta-heuristic is based only on local information; essentially no routing tables or other information blocks have to be transmitted to neighbor nodes in the network.

## 4 The Algorithm Design

In designing our algorithm for the wireless sensor networks, we have considered certain essential factors. The first is energy efficiency. Sensor nodes are most likely to be powered by batteries, and it is sometimes very difficult to change or recharge batteries for these nodes. Prolonging network lifetime for these nodes is a challenging issue. Another important factor is scalability and adaptivity to changes in network size, node density and topology. Some nodes may die over time, some new nodes may join later, some nodes may move to different locations and so on.

In this section we discuss the design of the Ant Intelligence Routing Algorithm (AIRA).

Before describing the components in AIRA, we first summarize our assumptions about the wireless sensor network and its applications as follows.

Sensor networks will consist of large numbers of nodes to take advantage of short-range, multihop communications to conserve energy. Most communications will occur between nodes as peers, rather than to a single base station. Nodes will be stationary or with limited mobility. Finally, in-network processing is necessary, implying that data will be processed as whole-message in a multi-hop fashion.

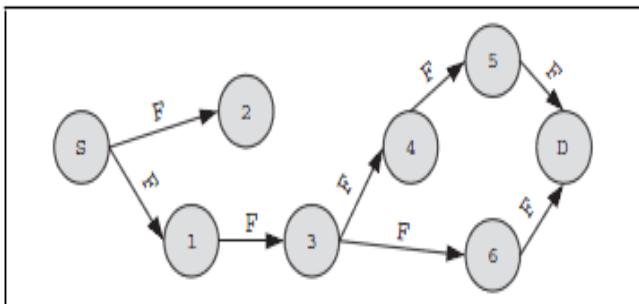
The routing algorithm is similarly constructed as many other ant colony based routing algorithms and consists of three phases.

### 4.1 Route Discovery Phase

This phase is responsible for generating new routes between source and destination. The creation of a new route requires the use of a forward ant and backward ant. A forward ant is an agent which establishes the pheromone track to the source node. On the other hand, a backward ant establishes pheromone track to the destination node. These two ants are similar in structure but differ in the type of work they perform. The forward ant is a small packet with a unique sequence number. Nodes are able to detect duplicate packets on the basis of the sequence number and the source address of the forward ant.

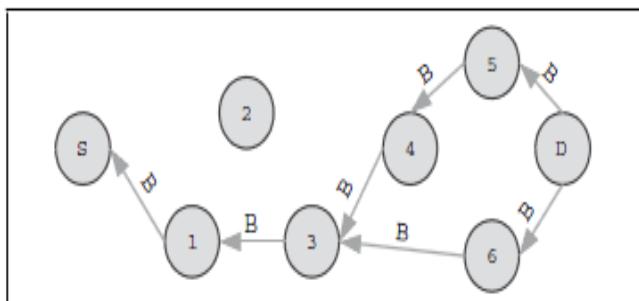
A forward ant is broadcast by the sender and relayed by the neighbors (intermediate nodes) of the sender until it reaches destination. A node receiving a forward ant for the first time creates a record in its routing table. The record in the routing table includes a destination address, next hop, pheromone value. The node interprets the source address of the forward ant as destination address, the address of the previous node as the next hop, and computes the pheromone value depending on the number of hops the forward ant needed to reach the node. Then the node relays the forward ant to its neighbors.

Duplicate forward ants are identified through their unique sequence number and destroyed by the nodes. When the forward ant reaches the destination node, its information is extracted and then it is destroyed. See Fig. 1.



**Fig. 1.** A forward ant (F) is sent by source (S) and relayed by intermediate nodes towards destination (D)

Subsequently, a backward ant is created with the same sequence number and sent towards the source node as in Fig. 2. The backward ant establishes a route to the source node. Once the source node receives a backward ant from the destination node, the path is established and data packets can be sent. Subsequent backward ants received are kept as alternate routes to the destination.



**Fig. 2.** A backward ant (B) sent by destination (D) and relayed by intermediate nodes towards source (S)

## 4.2 Route Maintenance

This phase is responsible for maintaining the route during communication. No special packets are needed for route maintenance. Once the forward ant and backward ant have established the route from the source to destination nodes, subsequently data packets are used to maintain the path.

When a node  $v_i$  relays a data packet toward the destination  $v_D$  to a neighbor node  $v_j$ , it increases the pheromone value of the entry  $(v_D, v_j, \phi)$  by  $\Delta\phi$ , thus the path to the destination is strengthened by the data packets. In addition, the next hop  $v_j$  increases the pheromone value of the entry  $(v_S, v_i, \phi)$  by  $\Delta\phi$ , thus the path to the source node is also strengthened. An acknowledgement is sent to every packet received. If an acknowledgement is not received within timeout period after a packet has been sent, then the route error message is transmitted to the previous node.

## 4.3 Route Failure Handling

The third phase handles route failures, which are caused especially through the death of nodes due to shortage of battery power and node mobility in the case of mobile sensors. Every packet is associated with an acknowledgment, hence a route failure is recognized through a missing acknowledgement. Upon detecting a route failure the current node sends a route error message to the previous node. If a node gets a route error message for a certain link, it deactivates this link by setting the pheromone value to zero.

Then the node informs its neighbors to relay the route error message towards the sender. On receiving the message, the sender searches for an alternative route in its routing table. If there exists a route, the sender sends the packet via this path. Otherwise, the sender initiates a new route discovery.

## 4.4 Adaptive Sleeping

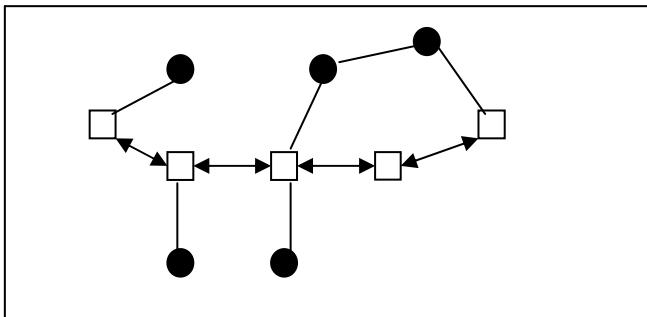
In many sensor network applications, nodes are idle for long time if no sensing event happens. Given the fact that nodes are battery powered and thus are power limited, it is not entirely necessary to keep nodes listening all the time. AIRA reduces idle listening time by putting nodes into periodic sleep. Additionally, unfortunately, in wireless networks, it is frequently the case that a packet transmission from one node to another will be overheard by all the neighbors of the sender within its transmission radius. All of these nodes will thus consume power needlessly.

Therefore in our protocol, nodes that may listen or interfere with a neighbor's communication are clocked to sleep when communication is in progress thereby avoiding overhearing and potential collision resulting in increased power efficiency.

There is a duration field in each transmitted packet that indicates how long the remaining transmission will be. If a node overhears a packet destined to another node, it knows how long to sleep by this field and sets a timer to wake it up later. During sleeping, the node turns off its radio.

The sleep interval can be changed according to different application requirements, which actually changes the duty cycle [7-13]. All nodes are free to choose their own sleep schedules when they are idle. However, to reduce control overhead, neighbor

nodes synchronize their schedules together. That is, any node after choosing its sleep schedule will broadcast to all its neighbors. If a node receives a neighbor's schedule broadcast before it chooses its own schedule, it follows the neighbor's schedule and goes to sleep synchronously with the neighbor. Nodes belonging to different schedules exchange their sleep schedules periodically ensuring that all neighbor nodes can communicate even if they have different schedules.



**Fig. 3.** Shows sleeping neighbors during data transmission

## 5 Experiment Results

In our experiments, we compared these protocols: IEEE 802.11, S-MAC and our algorithm. We included IEEE 802.11 because we considered it as a worst case scenario because of its high energy consumption due to idle listening. S-MAC was chosen, because it is considered energy efficient and is specifically designed for wireless sensor networks.

Our algorithm design and evaluation are for now based on simulation.

We implemented the algorithm in NS-2. We employed 50 sensor nodes according to IEEE 802.11 and covered a simulation area of 1000m x 600m. The simulation time varied according to the scenarios. Nodes were randomly positioned and were expected to have limited mobility.

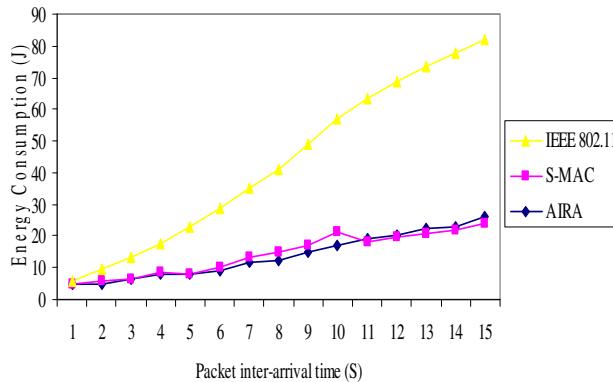
### 5.1 Energy Consumption

To measure the energy consumption on a radio, we measured the amount of time that the radio has spent in different modes: sleeping, listening, receiving or transmitting. The energy consumption in each mode is then calculated by multiplying the time with the estimated required power to operate the radio in that mode.

We vary the traffic load by changing the packet inter-arrival time of messages. If the packet inter-arrival time is 10s for instance, a message is generated every 10s. In this experiment, the packet inter-arrival time varies from 1 to 15s.

Fig. 1 shows the average energy consumption on radios in the entire network to transmit a fixed amount of data from the source to the destination. The result confirms

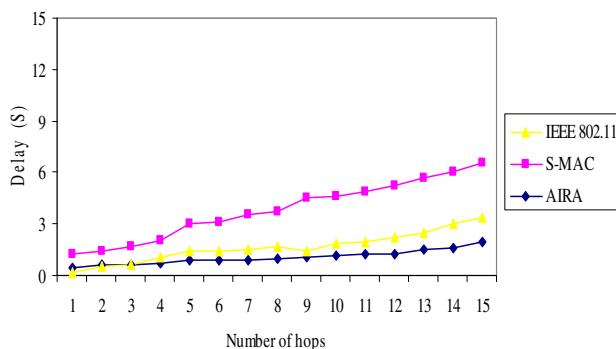
that AIRA achieves substantial energy savings over the IEEE 802.11 in multihop network, especially when traffic load is light. In addition, we can see that AIRA achieves better energy efficiency than S-MAC, especially when traffic load is heavy. The main reason is that S-MAC allows a node not to strictly follow its sleep schedule but to intermittently listen into a neighbor's communication to know whether current transmission is over or they are needed as a next hop before their sleep time is over. This increases the overall energy consumption in the network.



**Fig. 4.** Average energy consumption in the entire network

## 5.2 End-to-End Delay

In AIRA real-time delay is minimal because routes are established before transmission begins. Thus nodes in a selected route cannot sleep until transmission has ended and neighboring nodes that may interfere are clocked to sleep hence avoiding contention and collision resulting in greatly reduced delay. AIRA performs better than S-MAC and IEEE 802.11 in terms of end-to-end delay. Fig. 2 shows that AIRA has about the smallest average delay as compared to S-MAC and the IEEE 802.11. We also observed that delay increases with the number of hops.



**Fig. 5.** Average packet delay in a 15-hop network

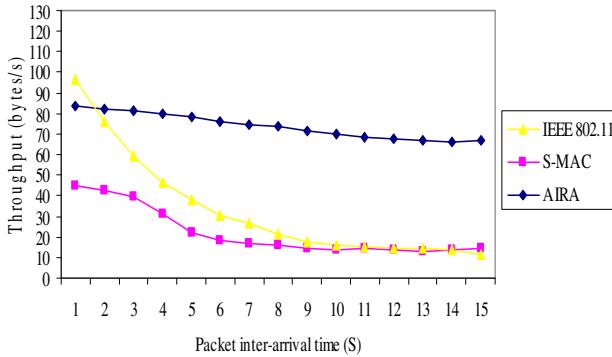
### 5.3 End-to-End Throughput

We considered throughput for a 15-hop network under varying traffic conditions to ascertain the maximum possible amount of data in a unit time. Data packets received perfectly at each hop were counted for the throughput and no error handling messages were counted as part of throughput.

Fig. 3 shows the throughput measured at each hop across the network under different traffic conditions. As expected, periodic sleeping during transmission reduced throughput in S-MAC. Throughput is lower because delay is higher.

The results also show that, for S-MAC and IEEE 802.11, throughput drops as the number of hops increases, due to the RTS/CTS contention in the multihop network.

AIRA achieves almost steady throughputs throughout the different traffic conditions due mainly to lack of contention and low delay.



**Fig. 6.** Throughput in a 15-hop network under varying traffic conditions

## 6 Conclusions and Future Work

This paper presents ant intelligence routing algorithm (AIRA), an ant colony based protocol specifically designed for wireless sensor networks that conserves power by turning off radios under certain conditions. Ant colony based algorithm is very promising because of its adaptivity, scalability and suitability to the dynamics of wireless sensor networks.

Energy efficiency is a primary goal in the protocol design. Low-duty-cycle operation of each node is achieved by adaptive sleeping. In addition, synchronization of individual sleep schedules enables AIRA to achieve significant energy savings in the entire network compared to IEEE 802.11. The protocol achieves significant power savings without compromising throughput.

We are working further to implement the protocol on real motes hardware. Our further investigations will include hotspot traffic behaviors, multimedia data handling and packet re-routing.

**Acknowledgment.** The first author would like to extend his profound gratitude to Professor Xiao Degui of college of computer and communication, Hunan University, for introducing him to swarm intelligence based routing.

The authors like to acknowledge the support of compatriots who listened to their presentations and offered suggestions. We like to specifically thank the members of the IOT group for their time and invaluable contributions.

## References

- [1] Sim, K.M., Sun, W.H.: Ant colony optimization for routing and load-balancing: survey and new directions. *IEEE Transn. on Systems Man and Cybernetics, Part A* 33(5), 560–572 (2003)
- [2] Di Caro, G., Dorigo, M.: Antnet: Distributed stigmergetic control for communications networks. *Journal of Artificial Intelligence Research*, 317–365 (1998)
- [3] Dorigo, M., Di Caro, G., Gambardella, L.M.: Ant algorithms for discrete optimization. *Artificial Life* 5(2), 137–172 (1999)
- [4] Wei, Y., Heidemann, J., Estrin, D.: Medium Access Control With Coordinated Adaptive Sleeping for Wireless Sensor Networks. *IEEE/ACM Transactions on Networking* 12(3) (June 2004)
- [5] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, IEEE Std. 802.11-1999
- [6] Stemm, M., Katz, R.H.: Measuring and reducing energy consumption of network interfaces in hand-held devices. *IEICE Trans. Commun.* E80-B(8), 1125–1131 (1997)
- [7] Kasten, O.: Energy consumption. Eldgenossische Technische Hochschule Zurich, [http://www.inf.ethz.ch/~kasten/research/bathtub/energy\\_consumption.html](http://www.inf.ethz.ch/~kasten/research/bathtub/energy_consumption.html)

# Research on Difference Distance Measurement and Localization Based on Zigbee

Zeng Gang

College of Information Science and Engineering  
Center South University, Changsha Hunan, China  
[zgise@mail.csu.edu.cn](mailto:zgise@mail.csu.edu.cn)

**Abstract.** According to research on range-based localization using wireless technology and focused on range attenuation characteristics of electromagnetic wave, a new method called difference distance measurement and localization is proposed to align the random errors. Results of the experiments using Zigbee-based hardware platforms indicated that precision of distance measurement and localization was improved.

**Keywords:** difference distance measurement, localization, Zigbee.

## 1 Introduction

There was lots of research on how to localize nodes in the Internet of things. As the basis of localizing, there are four classes of distance measurement method: distance measurement based on range attenuation characteristics of electromagnetic wave (RSSI), distance measurement based on electromagnetic wave propagation times (TOA), distance measurement based on the speed difference of electromagnetic wave and other waves (TDOA), and distance measurement based on the incidence angle of electromagnetic wave (AOA). With advantages of low cost and low complexity, the first class of distance measurement method is popular recently, for example, there are many implementations using Zigbee-based hardware platforms. According to the RSSI (Receive Signal Strength Indication), the propagation distance of radio signal is estimated in the receiving node, therefore with estimating distances from different radiate sources, the node is localized. Usually, distance estimating algorithms based on free-space propagation loss must be revised because of the propagation environment, for elements such as multipath fading always affect the range attenuation characteristics. For instance, the actual relationship between propagation distance and receive signal strength can be obtained by curve fitting using measured data [1][2][3], or the coordinates of measured node can be obtained with the vertical and horizontal offsets from coordinates of two reference nodes[4].

This article attempts to present an algorithm, with a difference method by using a pair of fixed reference nodes, the RSSI distance estimating result is revised, and the coordinates of measured node are obtained by using several pairs of fixed reference nodes. Furthermore, coordinates of measured node are estimated many times and averaged to be the final result. Experiment result indicated that precision of distance measurement and localization was improved.

## 2 Revised Algorithms Based on Free-Space Propagation Loss

Assume that reference nodes use omnidirectional antennas and radiate isotropic spherical wave, the transmit power is  $P_t$ , so the energy density at the distance of  $d$  is  $S = \frac{P_t}{4\pi d^2}$ , if the measured node receives the wave with omnidirectional antenna also and its orientation factor  $D = 1$ , its effective area  $A = \frac{D\lambda^2}{4\pi} = \frac{\lambda^2}{4\pi}$ , therefore the receive power of the measured node is

$$P_r = SA = P_t \left( \frac{\lambda}{4\pi d} \right)^2 \quad (1)$$

Thus,

$$\frac{P_t}{P_r} = \left( \frac{4\pi d}{\lambda} \right)^2 \quad (2)$$

After taking the logarithm on both sides of the equation (2), the power loss is

$$P_L = P_t - P_r = 20\lg 4\pi + 20\lg d + 20\lg f - 20\lg c \quad (3)$$

The equation (3) is the theoretical model based on free-space propagation loss, where  $f$  is frequency of the wave and  $c$  is speed of the light. Taking into account actual propagation environment, equation (3) is revised as the Shadowing Model, which is

$$\overline{[P_r]_{dbm}} = [P_0]_{dbm} - 10n\lg \left( \frac{d}{d_0} \right) + \xi_{dbm} \quad (4)$$

Where  $d_0$  is reference distance,  $P_0$  is the signal strength at the distance of  $d_0$ ,  $d$  is the actual propagation distance,  $\xi$  is the shadowing factor with unit of dbm and a Gaussian random variable with mean of 0,  $n$  is the path loss exponent whose value depends on the propagation environment. If simplify the equation (4) by ignoring the shadowing factor of  $\xi$ , the signal strength is given by

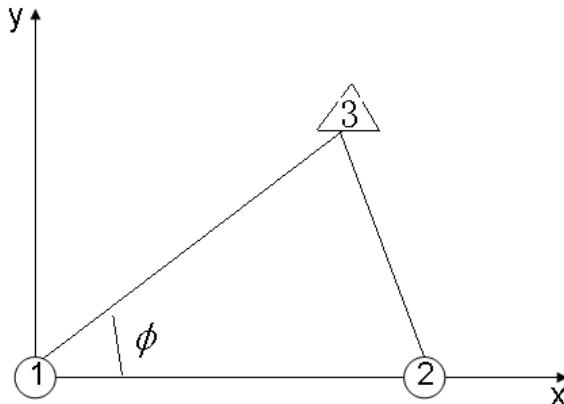
$$RSSI = \overline{[P_r]_{dbm}} = [P_0]_{dbm} - 10n\lg \left( \frac{d}{d_0} \right) \quad (5)$$

Where  $n$  is pending repeatedly measured.

## 3 Differential Distance Estimating and Localizing Algorithm Using Fixed Reference Nodes Pairs

According to equation (5), the value of RSSI always random jitters when being measured and expressed as the shadowing factor of  $\xi$ . If we obtained the distance of  $d$

from equation (5), clearly  $\xi$  should cause random deviation. For the purpose of revising this random deviation, difference distance estimating method is proposed in this article. In the chamber environment, if two receiving nodes measure the signal strengths as  $\overline{[P_{r1}]_{\text{dbm}}} = [P_0]_{\text{dbm}} - 10 \text{nlg} \left( \frac{d_1}{d_0} \right) + \xi_{1\text{dbm}}$  and  $\overline{[P_{r2}]_{\text{dbm}}} = [P_0]_{\text{dbm}} - 10 \text{nlg} \left( \frac{d_2}{d_0} \right) + \xi_{2\text{dbm}}$  from the same radiating source node at the same moment, we assume  $\xi_{1\text{dbm}} = \xi_{2\text{dbm}}$ , thereby the distance of  $d$  can be calculated by subtraction  $P_{r2}$  from  $P_{r1}$  instead of solving the equation (5). In this way, the random deviation caused by shadowing factor of  $\xi$  is reduced. We shall show the following algorithm.



**Fig. 1.** Cartesian coordinate system for differential localization

Shown in Fig.1, suppose nodes 1,2 are two fixed reference nodes, node 3 is what to be localized. Let node 1 be the coordinate origin, then establish the Cartesian coordinate system with connection between node 1 and 2 as the X-Axis. Thus

$$\overline{[P_{r12}]_{\text{dbm}}} = [P_0]_{\text{dbm}} - 10 \text{nlg} \left( \frac{d_{12}}{d_0} \right) + \xi_{12\text{dbm}}$$

$$\overline{[P_{r13}]_{\text{dbm}}} = [P_0]_{\text{dbm}} - 10 \text{nlg} \left( \frac{d_{13}}{d_0} \right) + \xi_{13\text{dbm}}$$

Where  $P_{r12}, P_{r13}$  are signal strength radiated by node 1 and received by node 2 and 3. Assume  $\xi_{12\text{dbm}} = \xi_{13\text{dbm}}$ , clearly

$$[P_{r13}]_{\text{dbm}} - [P_{r12}]_{\text{dbm}} = 10 \text{nlg} \left( \frac{d_{12}}{d_{13}} \right)$$

$$\frac{P_{r13}}{P_{r12}} = \left( \frac{d_{12}}{d_{13}} \right)^n$$

We have the relationship:

$$d_{13} = d_{12} \times \left( \frac{P_{r12}}{P_{r13}} \right)^{\frac{1}{n}} \quad (6)$$

Similarly,

$$d_{23} = d_{12} \times \left( \frac{P_{r21}}{P_{r23}} \right)^{\frac{1}{n}} \quad (7)$$

It is easy to show that

$$\cos \phi = (d_{12}^2 + d_{13}^2 - d_{23}^2) / (2 \times d_{12} \times d_{13})$$

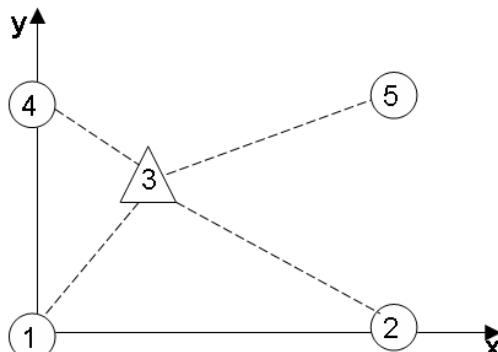
Finally we obtain the coordinates of node 3 on this Cartesian coordinate system are

$$\begin{cases} x = d_{13} \times \cos \phi \\ y = d_{13} \times \sin \phi \end{cases} \quad (8)$$

Furthermore, for the purpose of further revising the random deviation, shown in Fig.2, we deploy four fixed reference nodes 1,2,4,5 as four rectangle vertices. Let node 1 as the coordinate origin, then establish the Cartesian coordinate system with connection between node 1 and 2 as the X-Axis. If we let fixed node pairs (1,2), (1,4), (4,5), (2,5) be the reference node pairs, the application of equation (6), (7), (8) yields four pairs of coordinates (x1,y1), (x2,y2), (x3,y3), (x4,y4). Finally,

$$\begin{cases} x = \frac{1}{4} \sum_{i=1}^4 x_i \\ y = \frac{1}{4} \sum_{i=1}^4 y_i \end{cases} \quad (9)$$

are the localization result.



**Fig. 2.** Average localization result using several reference pairs

## 4 Chamber Experiment Result

The experiment environment was a chamber of eight meters by twelve meters rectangular area. Four fixed reference nodes were deployed as shown in Fig.2. Both the fixed nodes and measured node were Zigbee nodes with TI CC2430 chips. This chip supports radiation power programming and the power is hierarchical with eight levels and all the antennas are omnidirectional antennas. Fixed nodes package the RSSI value from measured node in IEEE 802.15.4 frames and send it to the measured node, programmer can access the RSSI value in register RSSI\_VAL of the CC2430 chip.

Intuitively, according to the RSSI value the solution of equation (5) should be revised with experience, in contrast with result yielded by application of equation (6),(7),(8). At the same time, the key to solve equation (6) and (7) is to determine the value of n by measuring, where we got n=3. To evaluate the experiment result, we randomly selected one hundred places in the experiment area and compared the estimated coordinates with actual coordinates, the deviation ranged from 1.6 to 2.5 meters, which indicated satisfactory accuracy.

## 5 Conclusion

Based on the free-space propagation loss and for the purpose of revise random deviation caused by propagation environment, the differential distance estimating and localization method is proposed in this article. Furthermore, the accuracy of measure is improved by averaging results using different fixed node pairs. In the experiment we got the accuracies of two meters on average, which was satisfactory.

## References

1. Zhang, H., Li, J., Lv, Y.: WSN Algorithm Based on Discrete Estimation. Computer Measurement & Control 17(1), 180–182 (2009)
2. Zhu, M., Zhang, H.: Wireless Location Technology for Indoor Environments Based on RSSI. Modern Electronics Technology 17, 45–48 (2010)
3. Wang, J., Wang, C.: Research of the Wireless Localization Algorithm Based on ZigBee. Instrument Technology (3), 44–46 (2010)
4. Li, L., Qin, G., Hu, N., Chen, K.: Wireless sensor node location approach based on transmission distance estimation. Systems Engineering and Electronics (1), 213–215 (2009)
5. TI/ Chipcon. CC2430 Preliminary Data Sheet (rev. 1.03) SWRS036A [DB/ OL ] :1632221

# Routing Optimization Based on Ant Colony Algorithm for Wireless Sensor Networks with Long-Chain Structure

Jing Gao<sup>1</sup>, Limin Wei<sup>2</sup>, Yongli Zhu<sup>1</sup>, and Lifen Li<sup>1</sup>

<sup>1</sup> School of Control & Computer Engineering  
North China Electric Power University  
Baoding Hebei, China

<sup>2</sup> Hebei Electric Power Design & Research Institute  
Shijiazhuang, China

**Abstract.** With the high demand of real-time and reliability, an ant colony algorithm with local search for lines monitoring sensor network is proposed. Heuristic function takes into account link delay, packet rate and distances from the sink node, local search is made every t iteration. Each node no need to maintain the global information, but needed to give the only Numbers. Simulation results show that the algorithm can quickly jump out of local optimum and find path with good performance in real-time and reliability.

**Keywords:** long-chain of wireless sensor networks, transmission line monitoring system, ant colony algorithm, routing optimization.

## 1 Introduction

Wireless sensor network (WSN) is a group of sensor nodes connected through wireless medium, it uses a large number of micro smart sensor nodes with AdHoc mode configuration, nodes work together to collect and process the information of the target network [1, 3]. It can be deployed in area of non-wiring, power supply difficulties and staff not easy to reach. The topology of monitoring system of power transmission lines is generally long-chain radial, the features of this multi-sensor network is centralized data receiving, multi-hop transmission, many-flow mode and so on [2].

Currently, the routing research of this WSN with long-chain structure is relatively small, the research of monitoring system of power transmission lines is even less. The routing algorithm of [7] is based on the fixed cluster head, cluster head is integration and forward the cluster data, but not for event-driven real-time applications. PEGASIS (Power-Efficient Gathering in Sensor Information Systems) protocol is a typical routing protocol with long-chain structure, the core idea of this algorithm is to use a greedy algorithm to generate a single chain with all nodes, chain nodes only communicate with their neighbors. The goal of this protocol is to balance node energy consumption and extend network lifetime, but when the node chain is too long, data transmission delay will be rapid growth, it is not suitable for real-time applications [6]. [9] proposed an ant colony optimization routing(ACO) algorithm with multi-cloud model, it uses multiple rules of cloud generator to adjust the pheromone residual factor,

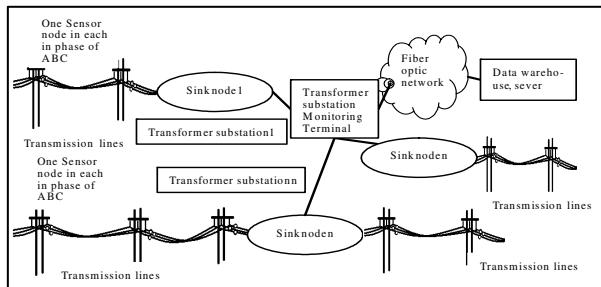
and using multiple ant colony to find multiple paths. [10]combination genetic algorithm and ACO, efficient but slightly complicated.

A lot of literatures tell us combination of local search and ACO is an effective way to obtain high-quality solutions. ACO shows a very rough search process, the solution can be locally optimized by local search. ACO with local search has been successfully applied to traveling salesman problem, in order of issue, machine learning and other issues. This paper, we proposed an ACO with local search for WSN routing problem of power line monitoring system. Local search is called every t generation, aimed at the optimal path as far. Simulation results showed that the algorithm can effectively jump out of local optimum and faster access to high-quality solutions.

## 2 About Transmission Line Monitoring Sensor Network

### 2.1 The Network Model of Transmission Line Monitoring System

The network structure of power line monitoring system shows in Figure 1. There are many towers with insulators on transmission lines, each phase of each tower is set on a sensor node. Sink node is to collect all the sensor nodes data and connected to server center. Node data collected via wireless links for multi-hop transmission, using Omni-directional antennas, and communication range can reach 2-3 jumps.



**Fig. 1.** Network structure of power line monitoring system

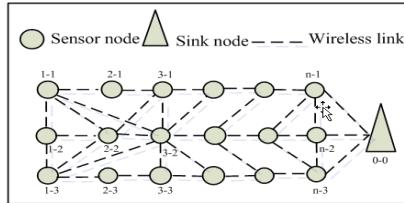
### 2.2 The Characteristics of WSN of Power Line Monitoring System

WSN of power line monitoring has the following characteristics: (1) node deployment relative rules and no longer move after deployment; (2) the node energy is not limited; (3) node localization and time synchronization can be resolved by installing GPS; (4) data transfer on real-time and reliability has high demands.

### 2.3 Node Number

The radial rule deployment brings convenience to the node number. Each node is given the only number. As this study only a line, so the number is a two-stage approach. Sink node numbered 0-0, the furthest layer away from the sink node are numbered 1-1,1-2,1-3.Previous paragraph of the number represents the distance from

the sink node, the greater the number, the closer the distance, the latter paragraph represents the position of the layer. Abstract model is shown in Figure 2.



**Fig. 2.** Model of wireless sensor networks of transmission line monitoring system

### 3 The Proposed of Line Monitoring Sensor Network Routing Optimization Algorithm

Social insects such as ants, although no vision, the colony was able to find the shortest path from food sources. People go through a lot of studies have found that ants during the exercise left a called pheromone substance on path, ant tend to move towards high strength of the material, exhibit a positive feedback: more ants walk on a path, the greater probability of selecting the path of the latter. Through this positive feedback, ant transfer from location i to location j, through a number of iterations the algorithm converge to the optimal results [4].

#### 3.1 The Process of Routing Establishing

##### 1) Preparation

- With GPS positioning system to obtain the location information of node and set number likes  $n_1-n_2$ ;
- Each node periodically broadcast their information by HELLO packets. A time stamp is included, communication delay can be obtained from the time difference of receive nodes and sending nodes;
- Calculate the link packet reception rate (PRR),  $PRR=L_r/L_s$ ,  $L_r$  is for the number of packets successfully received, and  $L_s$  is for the total number of packets sent;
- Set DE (Died End) is used to place the node of not feasible and congestion, initialized to empty. Reverse pressure beacon is sent to the child node when father node is overloaded, and put itself in the set DE. Set a timer for each node in the DE, node become active and participates in data forwarding again when the timer out.

Each node maintains a neighboring node table, use to record the number of neighbors, packet delay and PRR. Set the delay and PRR constrains:

$$delay(e_{c,v}) \leq D \quad (1)$$

$$PRR(e_{c,v}) \geq R \quad (2)$$

## 2) Building process

- a) Initial each edge of the network with 1 pheromone. Each nodes has a positive ant  $F_{s \rightarrow \text{sink}}$ , its task is to find a feasible path with high performance.
- b) In each node i,  $N_{c1}$  is the set of all neighbors to meet the constraints (1) and (2) and which  $n_1$  number is greater than node i;
- c) if  $|N_{c1}| > 0$ , then select the next traverse node based on transition probability  $P_{ij}$ ;

$$P_{ij} = \frac{[\tau_{ij}(t)]^\alpha \times [\eta_{ij}(t)]^\beta}{\sum_{h \in N_{c1}} [\tau_{ih}(t)]^\alpha \times [\eta_{ih}(t)]^\beta}, j \in N_{c1} \quad (3)$$

The formula of heuristic  $\eta_{ij}$  is as follows:

$$\eta_{ij} = \frac{PRR_{e<i,j>}}{\text{delay}_{e<i,j>}} + \frac{j_{n1} - i_{n1}}{r} \quad (4)$$

Heuristic take into account link quality, nodes delay and distance form sink node;

- d) if  $|N_{c1}|=0$ , that there is no node which closer to sink node than node i, then ants take a step back, and node i is placed into the set of DE;
- e) When reaches sink node, forward ant  $F_{s \rightarrow \text{sink}}$  will generate another backward ant  $B_{s \rightarrow \text{sink}}$ , it transferred all of its memories to  $B_{s \rightarrow \text{sink}}$ , and deleted itself. The path of  $B_{s \rightarrow \text{sink}}$  is exactly the same to  $F_{s \rightarrow \text{sink}}$ , but just the opposite direction. Forward and backward ants use a higher priority link queues than data packet, in order to accelerate the speed of path establishment. Backward ant is responsible for pheromone update, and the update rules are as follows:

$$\tau_{ij}(t+n) = (1-\rho)\tau_{ij}(t) + \sum_{k=1}^m \Delta\tau_{ij}^k(t) \quad (5)$$

$$\Delta\tau_{ij}^k(t) = \begin{cases} Q \times R^k, & \text{if ant- } k \text{ uses } (i, j) \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

Where  $R^k$  is the evaluation function of path p which is found by ant k, the smaller of the total delay and the higher of the PRR, the more pheromone can be growth. The growth of pheromone is proportional to the quality of the path.

$$R^k = \frac{PRR_p}{DELAY_p} \quad (7)$$

- f) Local search is made every t iteration; its object is the optimal path;
- g) End of the algorithm when the iterations reach the maximum iterations time.

## 3) Local Search Mechanism

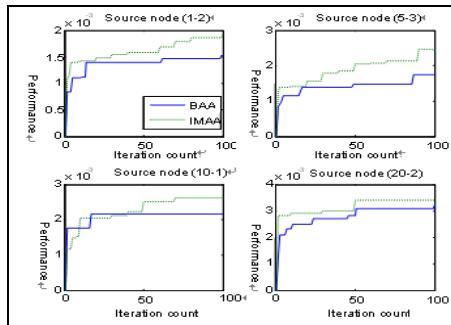
Consider the optimal path  $P_{s \rightarrow \text{sink}}$ , stop the local search as long as improvements. Local search steps are: select a node i in optimal path  $P_{s \rightarrow \text{sink}}$  by order, note the sub-path i to the sink node for the  $L_{i \rightarrow \text{sink}}$ , compare the merits of  $L_{i \rightarrow \text{sink}}$  and  $P_{i \rightarrow \text{sink}}$  according to the formula (7), if the  $P_{i \rightarrow \text{sink}}$  is fine than  $L_{i \rightarrow \text{sink}}$ , then replace the  $L_{i \rightarrow \text{sink}}$  with  $P_{i \rightarrow \text{sink}}$  in the path  $P_{s \rightarrow \text{sink}}$  and stop local search, otherwise select a next node for judgment.

### 3.2 Data Transmission

Data transmission is in the optimal path. When the node forwarding queue is too long, to ensure the real-time of forwarding, the node sends a warming message alert which records the node number  $n_1-n_2$ , and notified to other nodes by flooding the alert message, then place the node in the set DE. When the source node receives the alert message, it will look for a new child node other than  $n_1-n_2$ .

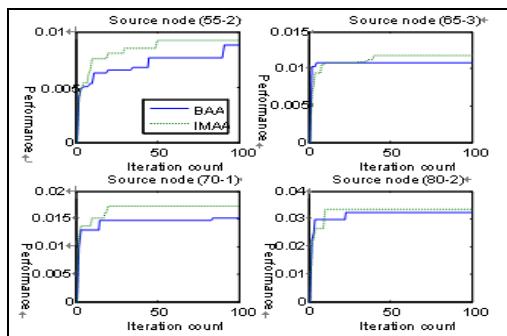
## 4 Simulation Results

Simulation platform is MATLAB 7.1, link delay and PRR is generated by system, delay in 1~10ms, PRR in 0.85~1, D=9.5ms, R=0.9. Based on the parameters of literature [5],  $\alpha=1$ ,  $\beta=4.5$ ,  $\rho=0.2$ ,  $Q=0.1$ , ants number is  $3*m$ ,  $r$  is the hops of communication,  $m$  is the number of layers of the long-chain networks.



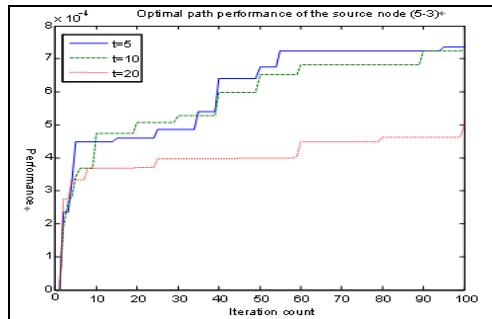
**Fig. 3.** Far layer nodes in optimal path performance changes

Figure 3 is in the case of  $m=100$ ,  $r=3$ ,  $t=10$ , performance evaluation is based on formula (7). Compared to the basic ant colony algorithm (BAA), the performance of path which is found by improved ant algorithm (IMAA) is much better, and can quickly jump out of local optimum to find the optimal path.



**Fig. 4.** Near layer nodes in optimal path performance changes

Figure 4 shows: the closer node layer from sink node, the result of application local search is not significantly as far-layers. Improved algorithm converges to the optimal solution before 50 times, after that local search does not improve the optimal solution.



**Fig. 5.** The convergence of improved algorithm under different t

Figure 5, reducing the value of t can speed up the speed of find optimal solution. In the case of t=5 and t=10, the optimal solution is much better than the case of t=20, but compare the case of t=5 and t=10, the optimal path is not much difference, the smaller of the value of t, the greater of the amount of computation.

## 5 Conclusions

With the high demand of real-time and reliability, an ant colony algorithm with local search for lines monitoring sensor network is proposed. Local search is made every t iteration for the optimal path with part greedy strategy. In addition, algorithm effectively avoids the loop by numbering each node. Reverse pressure beacon is used to relieve the pressure of a node when its task is too heavy, and prevent the occurrence of congestion. Simulation results showed that ant colony algorithm with local search can effectively jump out of the local optimization, and faster access to high quality solutions, so the algorithm is compared to the basic ant colony algorithm is more suitable for long-chain wireless sensor networks.

**Acknowledgments.** Our work is supported by the National Natural Science Foundation of china (NO.60974125). Most sincere gratitude to the National Natural Science Foundation of China.

## References

1. Tilakk, S., ABU-Ghazaleh, N.B., Heinzelman, W.: A Taxonomy of wireless micro sensor network models. *Mobile Computing and Communications Review* 1(2), 1–8 (2002)
2. Huang, X.: The design and application of insulator leakage current on-line monitoring system. *Electric Measurement & Instruments* 44(4), 9–14 (2007)

3. Sun, L.-M., Li, J., Chen, Y., et al.: Wireless sensor networks. Tsinghua University Press, Beijing (2005)
4. Dorigo, M., Stutzle, T., Zhang, J., Hu, X., Luo, X. (transl.): Ant Colony Optimization. Tsinghua University Press, Beijing (2007)
5. Feng, C.: Swarm Intelligence Optimization Algorithms and Their Applications. China University of Science and Technology (2009)
6. Yu, Y.-C., Wei, G.: An Improved PEGASIS Algorithm in Wireless Sensor Network. *Acta Electronica Sinica* 36(7), 1309–1313 (2008)
7. Wang, Y., Yin, X., You, D., et al.: A Real-Time Monitoring and Warning System for Electric Power Facilities Icing Disaster Based on Wireless Sensor Network. *Power System Technology* 33(7), 14–19 (2009)
8. Li, Z.: Local Search Strategy and Multiobjective Approach in Evolutionary Computation for Reactive Power Optimization. *Huazhong University of Science & Technology* (May 2010)
9. Li, L.-F., Zhu, Y.-L., Zhang, J.-Y.: A Cloud Model Based Multiple Ant Colony Algorithm for the Routing Optimization of WSN with a Long-Chain Structure. *Computer Engineering & Science* 32(11), 10–14 (2010)
10. Zhang, J., Zhu, Y., PengWei: Research on QoS Routing Optimization for Wireless Sensor Network with Large-scale Banded Structure. *Electric Power Science and Engineering* 26(4), 11–15 (2010)

# A Key Management of Wireless Sensor Networks for Telemedicine Care

Min Nan, Yong Xu, and Pengcheng Zhao

College of Mathematics and Computer Science  
Anhui Normal University  
Wuhu, China  
412216236@qq.com

**Abstract.** This paper introduces a key management of wireless sensor networks for hierarchy of telemedicine care, which is composed by q-Composite random key distribution and the idea of Hyperball-based key management. It makes full use of independence and association between collection sensors. Analysis shows this management has advantages in lower cost of calculation and communication, higher update efficiency, less storage and is applicable for telemedicine care networks.

**Keywords:** telemedicine, wireless sensor networks, key management.

## 1 Introduction

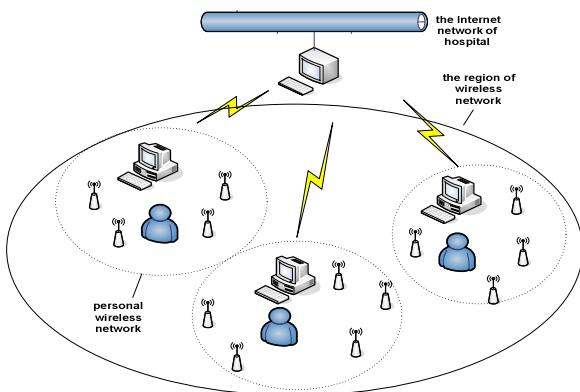
With the rapid development of wireless sensors and the Internet of things technology, the more miniature and more intelligence of biomedical sensors, the popularity of information process and data transmission by wireless, the development of intelligent wireless medical monitoring system has become a new generation of health care systems research. The doctors need medical information by network for the diagnosis, therefore, to solve the security problems of medical information's transmission has become a remote key to the quality of remote medical services and even to the development.

Remote medical monitoring system monitors the patient's body temperature, pulse, and respiratory rate through wireless sensor nodes real-timely, which are transmits from the collection sensors to the hospital's control center. The hospital's control center analyses the data and calls police if dates are abnormal. As the transmission of data related to patient's privacy, so safety is worth a high degree concerned for telemedicine services. However, key management is the basis for a variety of security issues, so the design of effective key management scheme is one of the most important issues we focus on. This paper presents a hybrid key management scheme under WSN, the scheme uses different methods of key management on different levels for the hierarchy of remote medical monitoring network. It not only takes advantage of independent and collaborative features between the sensors, which makes it have high key update efficiency and less stored keys, but also improves keys to have sufficient security between the personal wireless LAN networks and the

hospital's. Simulation results show that the scheme has a smaller overhead in calculation and communication, which can be applied to secure communications in wireless sensor networks.

## 2 Network Model

Remote medical monitoring network is a multi-level network architecture [1], which is by the personal base station equipment and bio-sensor nodes to form a dedicated micro-monitoring network. Sensor nodes need to use a central controller for monitoring of vital signs to control and collect data, and send data to personal base station equipment through wireless communication. The base station device integrates the data and transmits them to the network device, and the later transfers data to remote medical monitoring center through the Internet network. Professional medical person in hospital observe the data for statistics, provide the necessary advisory services and finish the remote medical. Remote medical monitoring network diagram is shown in Figure 1.



**Fig. 1.** Schematic diagram of remote medical monitoring network

According to the schematic diagram of the network, nodes in the network will be divided into three levels. The top is hospital monitoring center. And the second level is personal care base also called the cluster head node, which have strong computing power, more energy, enough storage space and large communication range. The cluster head is used to collect the data captured by sensors and transfer data to the external network. It is also responsible for issuing control information, directly or indirectly connecting hospital monitoring center. Capture terminal node which is also called the cluster member node is in the last level which has a large number, but their computing power, energy, storage and communication range are all limited. They are used to capture data in the monitoring area and to connect personal care base directly.

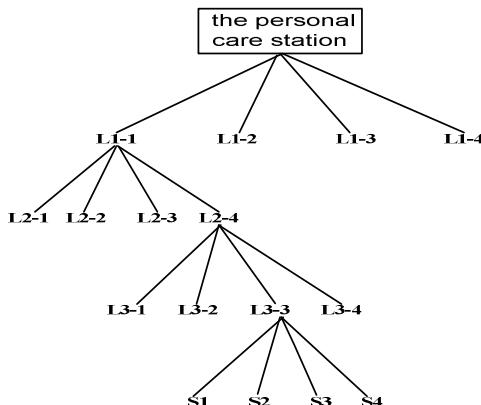
### 3 Key Management

#### 3.1 Solution Description

In view of the hierarchy of remote medical monitoring network, the key management program will also be divided into two levels. The one level is between the hospital control center and the personal care station, which uses q-Composite random key pre-distribution scheme [2], and the other level is between the personal care station and the capture sensors, also known as cluster head node and the cluster members, because the number of personal monitoring sensor nodes inside the region of wireless network are limited, we adopt the structure of hyperball-based key management scheme.

Paper [3] proposed the idea over hyperball, which is in the three-dimensional space any four non-coplanar points can decide a sphere, and the four points have the same distance to the center O of the sphere. When a point leaves and another point joins in, the new center of the sphere is calculated and the new sphere radius comes out. Therefore, even if the departure point knows the new center of the sphere coordinates, it still can't calculate a new radius. Apply this idea to key management, the storage of keys is not only less, but also can resist conspiracy attacks effectively.

Based on geographic location the capture sensors in personal wireless network regions are to be divided in groups of four. Coordinates of four points in each group may determine a sphere, and the center of the sphere is their parent node's coordinates. Similarly, the centers of the spheres also can group by four, and then constitute a number of spheres, and so on, until the root node. Any node changes will affect the center of the sphere which it belongs to and the center which it associates with until the root circle, the other centers are not disturbed. Each group based on hyperball can be abstracted to quad-tree form, shown in Figure 2.



**Fig. 2.** Schematic diagram of quadtree nodes

### 3.2 Key Pre-distribution

Before deploying keys hospital monitoring center need to generate a large key pool of M firstly, each cluster head node randomly selects the number of n ( $n << M$ ) different keys from the key pool. After deploying between the cluster head nodes need share at least q ( $q < n$ ) to establish a matching secret key. This can use one-way hash function, if the number of keys which two cluster head nodes share is t ( $t \geq q$ ), then the pairing key  $K = \text{hash}(k_1 || k_2 || \dots || k_t)$ .

Each sensor node randomly selects three different keys from sub-key pool of personal care base in the region to combine as the initial coordinate identification. The steps of key distribution for cluster member nodes are as follows, for node S4 as an example.

- Step1: Firstly S4 need to pass authentication through the cluster head node, and then S4 randomly selects three keys from the sub-key pool to generate a coordinate and sends it to the cluster head.
- Step2: The cluster head adds S4 into the L3-3 group, and then publicly sends a coordinate point X to S4. S4 receives the packet and finds the distance R between the coordinate X and its own coordinate, that R is the group key.
- Step3: Cluster head updates the coordinates of L3-3, L2-4 and L1-1. L3-3 for example, cluster head calculates the point P' according to S1, S2, S3, so that the distance between P' and 3-point are all R. Then P' and the new sphere center coordinates L3-3' encrypted using R for S1 to S4 are sent publicly. D1, D2, D3 use P' and their own coordinates to calculate R after receiving, then decrypt L3-3' with R. While S4 receives the packet and uses the known R can directly derive L3-3'. The update method of L2-4 and L1-1 are on the analogy of this.
- Step4: Update the cluster head. Calculate point P based on L1-2, L1-3 and L1-4, the distance between the coordinates of point P and the 3-point is R. Then sent point P publicly. After L1-2, L1-3 and L1-4 receiving, calculate R through the coordinate P and their own coordinates, the R is the new group key.

### 3.3 Key Update

In order to higher level of security, we can use a periodically update way. According to the patient's degree of attention to privacy, we set the update period T to replace the keys inside the sub-key pool of personal care base station, and re-select keys from the large key pool of the hospital monitoring center. At the same time, the coordinates' identification of end sensors should be initialized once every T cycles.

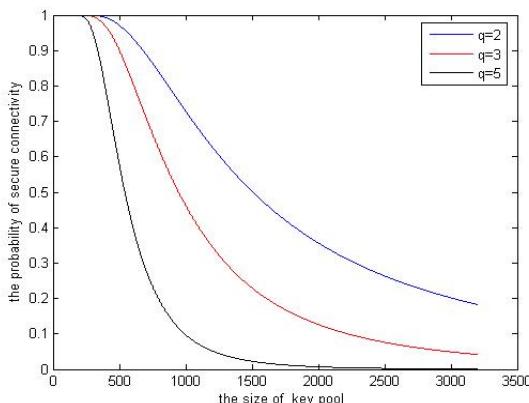
## 4 Safety Analysis

In this scheme of the three layers network architecture all the cluster members are directly connected with the cluster head, while cluster head directly or indirectly connected to the hospital monitoring center. Therefore, security connectivity of the

first layer affects the entire network security connectivity, so we only need to consider the connectivity on cluster head level.

The method of selecting  $n$  kinds of keys from the key pool of the hospital monitoring center for cluster head node are  $C_M^n$ . So a total of the two cluster heads select different  $n$ -keys approach are  $C_M^n \cdot C_M^n$ . Assuming the probability of the two clusters share  $i$  keys is  $P(i)$ , then:  $P(i) = \frac{C_M^i \cdot C_{M-i}^{2(n-i)} \cdot C_{2(n-i)}^{n-i}}{C_M^n \cdot C_M^n}$ .

Therefore, according to total probability formula, the probability of sharing at least  $q$  keys between two cluster head is  $P = 1 - \sum_{i=0}^{q-1} P(i)$ . Specific relationship is shown in Figure 3, where  $n$  is taken 50.



**Fig. 3.** Secure connectivity diagram

On another layer between cluster head and cluster members the key management program assumes the total numbers of end sensor nodes are  $N$ , then the complexity for the algorithm update of Hyperball idea is  $O(\log N)$ . Each cluster member has  $\log_4 N + 1$  keys, the total numbers of keys which cluster head manages are  $(4\log_4 N - 1)/3 = (2\log_2 N - 1)/3$ . At the same time, the scheme also meets the demands for wireless sensor network security in the basic communication, which are the group key secrecy, forward secrecy and backward secrecy. As the communication keys between the cluster members and cluster head are independent of each other, even if one or some of the cluster members are captured, this will not reveal key information of secure communication between the other nodes.

In summary, the proposed key management scheme has good security, and the cost is moderate for telemedicine monitoring network.

## 5 Conclusion

This paper presents a key management scheme of wireless sensor networks for hierarchy of telemedicine care, which has a small computational overhead and higher communication security. It can be applied to wireless sensor network environment which has limited computing power, and in the transmission can effectively protect the privacy of patient's. So it can be applied to remote medical monitoring network. Next we will further improve the key management scheme of the network of remote medical monitoring, and begin to study key management scheme for remote medical consultation network.

**Acknowledgments.** This work was supported by the National Science Foundation of Anhui Province under grant 11040606M137.

## References

1. Zhao, Z., Cui, L.: A Remote Health Care System Based on Wireless Sensor Networks. *Information and Control* 35(2), 265–269 (2006)
2. Chan, H., Perrig, A., Song, D.: Random key predistribution schemes for sensor networks. In: Proc. of the 2003 IEEE Symp. On Security and Privacy, pp. 197–213. IEEE Computer Society, Washington (2003)
3. Chen, S.J., Feng, N.R., Xu, Y., Li, J.: A new method of secure multicast rekeying based on hyperball. *Journal of Anhui University of Technology and Science* 22(3), 48–50 (2007)

# An Improved Routing Algorithm on LEACH by Combining Node Degree and Residual Energy for WSNs

Weiping Luan, Changhua Zhu, Bo Su, and Changxing Pei

State Key Lab of Integrated Services Networks, Xidian University

Xi'an, 710071, China

qingqingleizhu@163.com, {chhzhu, chxpei}@xidian.edu.cn,

subwurlin@hotmail.com

**Abstract.** According to the limitation of battery capacity and memory size for node of WSNs (Wireless Sensor Networks), the design of efficient routing protocol is one of the most critical issues for reducing energy consumption and prolonging the lifetime of the network. In this paper, we analyze LEACH (Low-Energy Adaptive clustering Hierarchy), a clustering-based protocol and its limitations, and introduce a new weight defined by combining node degree and residual energy. We propose an improved algorithm based on the weight. Simulation results show that the improved algorithm NDEA (Node Degree and Energy-Aware routing protocol) can optimize clustering and balance network load. Furthermore, it can greatly improve the network lifetime.

**Keywords:** routing protocol, node degree, residual energy, LEACH, WSN.

## 1 Introduction

With the development of the advancement of wireless communication, micro-electronics technology, computer network and sensor technology, the Wireless Sensor Networks (WSNs) have aroused more and more attention. The network is composed of lots of inexpensive micro-sensor nodes that form a multi-hop self-organizing network system through wireless connection. The information of objects in the monitoring area can be obtained through perception, collection and processing cooperatively[1], then the valid data will be sent to observers by sink via satellites or Internet.

Sensor nodes integrate vast functions in a tiny volume such as information gathering, data processing and wireless communication, and they are usually powered by batteries with limited energy. Hence, environmental factors and energy drain of nodes may cause changes of network topology, or even paralyze the whole network; besides, sensor nodes are distributed in a wide and complex area. In many cases, some nodes are so harsh that staff members cannot reach them, so it is not realistic to provide supplementary energy by battery replacement[1,2]. Therefore, how to maximize the life span of network with limited energy is the biggest challenge for sensor networks.

Routing protocols in conventional wired networks and ad-hoc networks cannot be used in wireless sensor network because the energy limits. As a result, the research

based on an energy-efficient routing protocol for wireless sensor networks is particularly important. In this paper, we will analyze the basic ideas of LEACH (Low-Energy Adaptive clustering Hierarchy) protocol and its limitations, and propose an improved mechanism, using a weight combining neighbor nodes degree and residual energy. The results of simulation by NS2 have proved that the improved algorithm NDEA (Node Degree and Energy-Aware routing protocol) can optimize clustering, balance network energy and prolong lifetime of the network.

## 2 LEACH Protocol and Its Limitations

### 2.1 Basic Ideas of LEACH Protocol

LEACH is the first hierarchical routing protocol that presents data aggregation, which is also known as data fusion. In the LEACH protocol, channel is shared through TDMA. The routing protocol is a self-organizing, adaptive clustering protocol that uses randomization to distribute the energy load evenly among the sensors in the network. In LEACH, the nodes organize themselves into local clusters, with one node acting as the local base station or cluster-head[2]. In addition, the operation of LEACH is broken up into several rounds, and each round consists of three phases including cluster-head selection, cluster set-up and data transmission. The cluster-heads of this round are elected in the first phase. The main selecting principle is: each sensor node generates a random number between 0 and 1, if the random number is less than the threshold  $T(r)$  which has been defined in advance, it will be selected as the cluster-head; if not, it will declare itself the member node. It's noteworthy that each node can act as cluster-head just once.

$$T(r) = \begin{cases} \frac{p}{r \bmod \frac{1}{p}} & n \in G \\ 1 - p \times \left( r \bmod \frac{1}{p} \right) & \text{otherwise} \\ 0 & \end{cases} \quad (1)$$

Where  $p$  is the desired percentage of cluster heads,  $r$  is the current round, and  $G$  is the set of nodes that have not been selected as cluster-heads in the last  $1/p$  rounds[2].

In the second phase, each node that has selected itself as a cluster-head for the current round broadcasts an advertisement message to the rest of the nodes. In this phase, the cluster-head use a CSMA MAC protocol, and all cluster-heads transmit their advertisement using the same transmit energy. The non-cluster-head nodes must keep their receivers on during the set-up phase to hear the advertisements of all the cluster-head nodes[2]. In this phase, each non-cluster-head node decides which cluster it will belong to for this round. The decision is based on the received signal strength of the advertisement. After each node finishes the decision, it must inform the cluster-head that it will be a member of the cluster[2,3,4].

After the cluster-head nodes receive all the messages for nodes that will be included in the cluster, it creates a TDMA schedule, telling each node in the same cluster when it can transmit data. This schedule is broadcasted back to the nodes in

the cluster[2,5]. At last, the data transmission phase begins and cluster-head nodes send collected data to sink.

## 2.2 Limitations of LEACH Protocol

LEACH protocol reduces the energy loss effectively in communication by randomized rotation in which cluster-head node rotates among the various sensors in order to distribute the energy load evenly and prolong the lifetime of network. In addition, LEACH performs local data fusion to “compress” the amount of data sent from the clusters to the base station, which further reduces the data redundancy. The use of TDMA/CDMA reduces the conflict within cluster and that between clusters. This hierarchical routing protocol is a breakthrough compared to the flat ones. The lifetime of network is extended by trimming part of energy. However, there are still some limitations in LEACH:

1. The selection of cluster-head is determined by random system without taking the residual energy of nodes into account, so once a node with less energy is selected to be cluster-head, its energy will dissipate quickly, and then nodes in this cluster will lose touch with its cluster-head and sink. In other words, it will accelerate energy dissipation, shorten the lifetime of the network, and may put the network out of action at worst[4,5].
2. LEACH didn't specify how the cluster-head nodes are located in the network. The simulation results confirm that all the cluster-heads may be concentrated in a certain part of the network, and none will appear in other parts.
3. Once cluster-head nodes are determined, they need send advertisement message to all the other nodes in LEACH. That means a large transmitter power of cluster-head nodes is required, hence LEACH does not apply to large-scale networks.

## 3 An Improved Algorithm NDEA

Based on the limitation described above that LEACH does not take remaining energy and load balancing into account, we introduce a new weight which is a function of neighbor nodes degree and remaining energy to choose the cluster-head nodes.

### 3.1 Basic Idea of NDEA

First of all, we define a weight  $W(n)$  for each node. The value of  $W(n)$  is the sum of weighted neighbor nodes degree and residual energy[2].

$$W(n) = w_1 f(d_n) + w_2 g(e_n) \quad (2)$$

Where  $w_1, w_2$  is the weights, they can be changed for different applications.  $f(d_n)$  is a function of neighbor nodes degree, and  $d_n$  is the neighbor nodes degree of node  $n$ ,  $g(e_n)$  is a function of residual energy, and  $e_n$  is the residual energy of node  $n$ . Here we define  $f(d_n)=d_n$ ,  $g(e_n)=e^{c_n}$ , the node with biggest weight in its region is considered to be the candidate of cluster-head node.

In NDEA a energy threshold  $\min(E)$  is also set for cluster-head selection. when a node has maximum weight and meanwhile its residual energy is greater than  $\min(E)$ , it could be a candidate of cluster-head node, otherwise, it can't be. Its main working process is as follows:

Step1: Collection of node information. The information of neighbor nodes degree and residual energy should be prepared first for weight of nodes. In this phase, each node broadcasts *helloID\_msg* to its neighbor nodes. The message contains the ID of the node. Meanwhile, at each node an ID table is built to hold the received message. The value of weight can be calculated by residual energy and the information of neighbors and can be added to *helloworld\_weight\_msg*, then each node broadcasts *helloworld\_weight\_msg* to its neighbors. In the next rounds, nodes need to broadcast *helloworld\_weight\_msg* only. However, when a node is about to run out of energy, it will broadcast *hellodead\_msg* to its neighbors, and this node will be removed from their ID tables of its neighbors.

Step2: Selection of cluster-head nodes. If the residual energy of a node is less than  $\min(E)$ , the node will declare to be the member node; if it is greater than  $\min(E)$ , we will compare weight of this node with that of its neighbors, if the weight of this node is the biggest, it will be declared to be the cluster-head; if the weight of this node is equal to that of one or more neighbors, the node with less ID will declare itself to be cluster-head; the nodes that fail to be cluster-head will wait for the cluster-head advertisement message. If some of them do not receive the message after a certain period of time, they will send *cluster\_head\_msg* to the node with maximum weight in its neighborhood, and wait for cluster-head advertisement. Those nodes that are not cluster-heads declare to be member nodes.

Step3: Cluster set-up. Those nodes that are decided to be the cluster-head should broadcast cluster-head advertisement to their neighbors. After this phase completes, each non-cluster-head node determines which the cluster it will belong to for this round. This decision is based on the received signal strength of the advertisement.

Step4: Data transmission phase. After clusters are built, the cluster-head node creates a TDMA schedule telling each node when it can transmit. Once the TDMA schedule is fixed, data transmission will begin[2].

## 4 Experiment Platform and Simulation Comparision

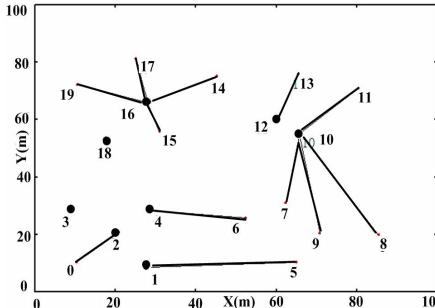
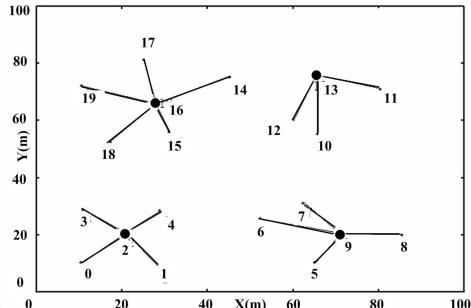
In order to evaluate the performance of NDEA algorithm, we extend the network simulation software NS2 to simulate NDEA and LEACH. This will give us a clearer picture of advantages and disadvantages of the different protocols. Simulation parameters are shown in TABLE I.

Assume that the lifetime of network is the duration from the start to the time when only 20% of nodes are alive.

The validity of the algorithm is measured by topology, network lifetime and energy consumption. In Fig. 1 the clusters created by LEACH are shown when round=1, node 1, 2, 3, 4 are all cluster-head nodes, and they are so concentrated that there is no node in their clusters. Meanwhile, node 3 and node 8 are selected to be the cluster-head, but no nodes in their clusters except themselves.

**Table 1.** Simulation Parameters

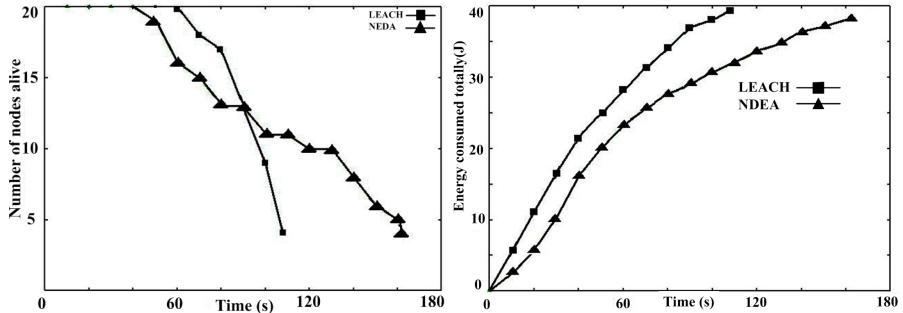
parameter	value	parameter	value
Size of network	100(m)100(m)	Transmission range	changeable
Location of sink	(50,50)	Propagation model	TwoRayGround
Number of nodes	20	Frequency	914e+06
Initial energy	2J	$w_1, w_2$	0.5, 0.5

**Fig. 1.** Clusters of LEACH(round=1)**Fig. 2.** Clusters of NDEA (round=1)

The clusters formed by NDEA are shown in Fig. 2 when round=1. The clusters with consideration of load balancing and residual energy are more reasonable compared to those of LEACH.

In Fig. 3 the variation of the nodes alive over time in LEACH and NDEA is shown. Obviously, the lifetime of the network with NDEA is longer than that with LEACH. But in NDEA, the death of the first node occurs at 40s, earlier than that in LEACH. It is due to the broadcast during network initialization.

In Fig. 4 the variation of energy consumed over time is shown. Plainly, total energy consumption of the network with NDEA is less than that with LEACH at any time. And Fig. 5 shows the amount of data sent to BS over time.

**Fig. 3.** Changes of nodes alive over time**Fig. 4.** Changes of energy consumed over time

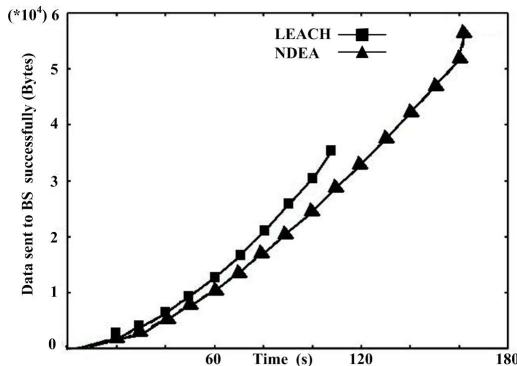


Fig. 5. Data sent to BS over time

## 5 Conclusion

In this paper, we analyze the LEACH protocol and propose a new algorithm NDEA, which reduces global energy usage by distributing the load to all nodes and considering its residual energy. However, there are still some limitations in NDEA. First of all, the decision of cluster-head depends on its neighbors, requiring several comparisons. Therefore, the convergence of the network is slow; secondly, neighbor information collection is needed during the network initialization, which caused additional overhead.

## References

1. Sun, L.M., Li, J.Z.: Wireless Sensor Networks. Tsinghua University Publishing House, Beijing (2005) (in Chinese)
2. Heinzelman, W.R., Chandrakasan, A., Balakrishnan, H.: Energy-Efficient communication protocol for wireless microsensor networks. In: Proc. of the 33rd Annual Hawaii Int'l Conf. on System Sciences, pp. 3005–3014. IEEE Computer Society Press, Maui (2000)
3. Hwang, K.S., Wu, C.S., Cheng, B.C.: An improved LEACH-based power aware routing protocol in wireless sensor networks. In: Proc. The Third International Conference on Communications and Networking in China, Beijing, China, pp. 726–731. IEEE (2008)
4. Hu, J., Shen, L.F., Song, T.C., et al.: Novel clustering algorithm for wireless sensor networks. Journal of Communications 29(7), 20–26 (2008)
5. Deng, Z.X., Qi, B.S.: Three-layered routing protocol for WSN based on LEACH algorithm. In: Proc. Wireless, Mobile and Sensor Networks, Shanghai, China, pp. 72–75. IET press (2007)

# Priority-Based Random Access Algorithm for TD-SCDMA Trunking System\*

Qing Jiang, Xianglin Wu, Hao Chen, and Xueqian Wang

Chongqing Key Lab of Mobile Communications Technology  
Chongqing University of Posts and Telecommunications (CQUPT)  
Chongqing, China  
jiangq@cqupt.edu.cn, 693421947@qq.com

**Abstract.** In the TD-SCDMA trunking communication system, users need to have predefined priority in the trunking group, according to the users priority access the uplink traffic channel, the users with the high priority give priority to make use of these channels. To solve these problems, we proposed one random access algorithm based on priority to maximize the access success rate and provide the high priority users with better QoS (Quality of Service). For further discussion, three kinds of algorithms were given to divide the priorities in this article. Each algorithm was simulated and its applicable situation was analyzed.

**Keywords:** TD-SCDMA, trunking system, random access, QoS, priority.

## 1 Introduction

In more recent years, trunking communication system has developed, which is a kind of more economic and more flexible scheduling communication system. At the same time, it's widely used in our daily life. Such as the government, the energy and the transportation, the airport and the shipside, the water conservancy and military. It's used to meet the needs of the communication within an organization. The present trunking communication system developed based on the second generation mobile communication technology. Namely, GSM(Global System for Mobile) technology integrates the technology of trunking communication system into public mobile communication network. Considering China's third generation mobile communication standard, named TD-SCDMA system, which is mainly based on our intellectual property rights and widely accepted and recognized by the international, but there have many unique advantages in some ways. It's necessary for us to study the standard of TD-SCDMA in the trunking communication.

In the trunking communication system, when multiple users simultaneously occupied the same uplink traffic channel, they will conflict, which leads to obstruction so that they would affect network performance. When different priority users

---

\* Supported by National Science & Technology Major Program (2012ZX03004009), the special fund of Chongqing key laboratory (CSTC) and the transformation project of excellent achievement of Chongqing Municipal Education Commission (Kjzh11206).

conflicted, in order to guarantee the QoS of the network, we should ensure the access successful with the high priority users and block the low priority users. So, it is necessary to analyze the division of priority in the multiple priority trunking communication system. It can provide certain theoretical basis to the actual construction of TD-SCDMA trunking communication system.

## 2 Introduction the TD-SCDMA Trunking Communication

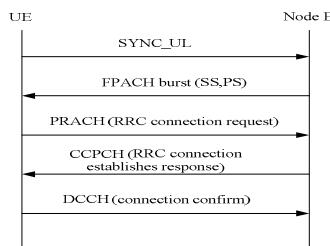
Providing the professional scheduling function is a technology development direction in the public mobile communication system. The mobile communication system, supporting trunking business is a special scheduling communication. It can make multiple users automatically share a small number wireless channel, and trunking users use these channels dynamically.

Digital trunking communication based on TD-SCDMA modifies related terminal, wireless access network and core network, under the condition that maintaining the integrity of current 3G access network. Defining a special trunking server to achieve trunking function based on TD-SCDMA public system network structure, and adding the interface that support trunking and related protocols[1]. The whole network includes three parts: user terminals, wireless access network and core network. Besides providing common voice communication, TD-SCDMA trunking communication requires to provide multi-media services and transmitting image, even real-time transmitting video, etc.

### **3 Random Access Algorithm Based on Priority for TD-SCDMA Trunking System**

### 3.1 Random Access Process of TD-SCDMA System

Random access is a process of mobile terminal requesting access to the system. Random access can be used not only to the first mobile access, supporting launch call, but also the registration, paging response, short message and sudden type of data transmission [2], etc. Fig.1 briefly describes the process of setting up call in TD-SCDMA.



**Fig. 1.** Random access process

UE send SYNC-UL (Uplink Synchronization Code) to base station at initial transmit power. If the base station receives SYNC-UL successfully, it will send FPACH (Fast Physical Access Channel) to UE as a response. If UE don't receive the confirmation message in four subframes after sending SYNC-UL, it will resend SYNC-UL to base station after a random time delay. When UE has sent SYNC-UL Max times and didn't receive the confirmation message from the base station, it will begin the next access process [3,4]. Random access is a process of mutual delivering message between UE and base station.

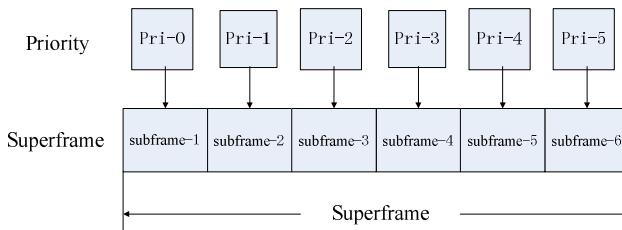
### 3.2 Random Access Algorithm Based on Priority

In order to make TD-SCDMA mobile communication system support trunking services, it must make the downlink traffic channel share, and the uplink traffic channel adopts time division multiplex through preempting of the trunking users in the system, thus achieves the uplink communication function, as follows.

- 1) The users who have the same PPT (Push-To-Talk) authority in the trunking group send uplink access request message to TSC (Trunking Service Controller) or DC (Dispatch Console) through BSS (Base Station Subsystem).
- 2) TSC or DC decides which trunking user can use the uplink traffic channel.
- 3) TSC or DC send confirmation message of access uplink traffic channel and forbid the access to downlink traffic channel. So as to stop other PPT trunking users send access message repeatedly until the uplink communication is over.

There are some disadvantages in this process. The trunking users' priority and authority are same and there is no pre-priority in the group. So the trunking users can't access the uplink traffic channel according to the priority and the system can't achieve the function that high priority user can use the uplink traffic channel firstly.

We can advance to set priority for the users within the group in trunking system[5], for example this paper set up six priority for the users within the group in trunking system, some sub-frame will be part of a super-frame, when the user need send random access information, let different priority users send SYNC\_UL in the corresponding sub-frame of super-frame, as shown in Fig.2.



**Fig. 2.** The mapping relationship of priority to subframe

Among them, “Pri-0” is the highest priority users. “Pri-5” is the lowest priority users. When the user send random access information, and it will send SYNC-UL in the corresponding sub-frame according to the priority of the users. Namely, the user of

“ $Pri\_i$ ” priority send SYNC-UL in the corresponding “ $i+1$ ” sub-frame. So it reduced the collision probability of users effectively, and it’s counseling the congestion condition of system when the user access strength is big, it improved the access success rate of users and greatly improved the QoS of the system[6].

### 3.3 Three Algorithms of Dividing Priority

We introduce the priority distinguish based on random access algorithm, Users will be divided into several different priority level, and the different priority users access the corresponding different sub-timeslot frames, this paper uses three different functions to divide the priority, and analyzed the corresponding environment of three algorithms through comparing each algorithm.

#### 1) Linear function algorithm

The use of specific formula as follows:

$$G[n] = a + k \cdot n. \quad (1)$$

Among them,  $n=0:5$  (the number of priority),  $G[n]$  is the number of n-priority users. If we set specific parameters, which are  $a$  and  $k$ , then we can get the number of corresponding priority users.

#### 2) Exponential function algorithm

The use of specific formula as follows:

$$G[n] = a + b \cdot k^n. \quad (2)$$

Among them,  $n=0:5$ (the number of priority),  $G[n]$  is the number of n-priority users. If we set specific parameters, which are  $a$ ,  $k$  and  $b$ , then we can get the number of corresponding priority users.

#### 3) Logarithmic function algorithm

The use of specific formula as follows:

$$G[n] = a + \log(b + n) / \log(k). \quad (3)$$

Among them,  $n=0:5$ (the number of priority),  $G[n]$  is the number of n-priority users. If we set specific parameters, which are  $a$ ,  $k$  and  $b$ , then we can get the number of corresponding priority users.

We can know various suitable environment of three algorithms through analyzing the success rate and time delay of the users' accession. Based on these, we would use different dividing priority algorithms in corresponding special trunking system.

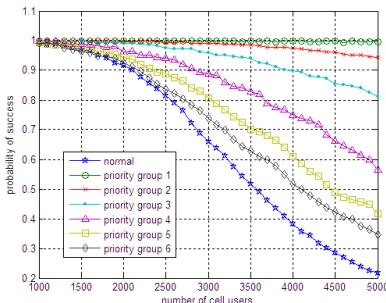
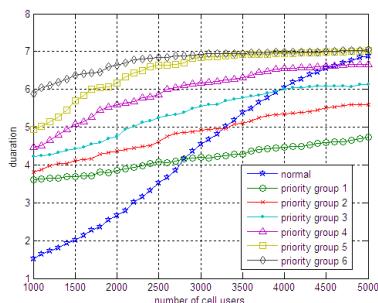
## 4 Simulation

The random access algorithm based on priority this paper proposed has been simulated on matlab software platform. The performance of simulations results are showed in followed figures. The simulations parameters we set as shown in table.1.

**Table 1.** The simulation parameters of Random access process

parameter	value
average user call per second	0.3
the biggest times of retransmission	4
the length of sub-frame	5ms
the number of simulation sub-frame	30
the length of simulation	150ms
the loop times	1000

We verify the performance of the random access algorithm based on the priority. When the number of group users gradually increases in the cell, we judge the performance of the algorithm, through analyzed. Fig.3 and Fig.4 show the comparison of the success rate and time delay of the users' accession in the simulation environment, respectively based on the priority algorithm and the general algorithm. Where, “normol” is the general algorithm; “priority group  $i$  ( $i=1,\dots,6$ )” are the different priority user in the priority algorithm; “priority group 1” is the user whose priority is highest

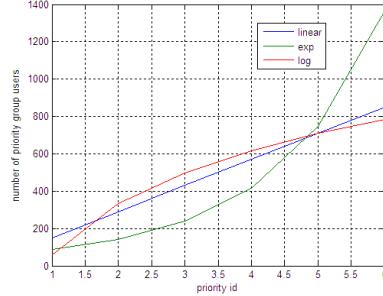
**Fig. 3.** Comparison of access success rate in two cases**Fig. 4.** Comparison of the access time delay in two cases

Based on Fig.3, the success rate of user's access in priority-based random access algorithm is significantly higher than the general algorithm, especially in the high load condition. The fact that users access rate is significantly higher than the general algorithm, can greatly improved the QoS of the system. By priority from low to high, their access success rates increase in turn, which acquires obvious effect. Under condition that access capacity is great, high priority users can smoothly access, which makes a high QoS support for these users. Meanwhile, low priority users acquire higher access rate than the general algorithm.

In addition, based on Fig.4, priority-based random access algorithm, as users wait for time slot to access, in condition that the user' capability is small, the access delay is higher than in the general algorithm, but as the group number of users increases, the probability of users' collision in the general algorithm becomes large, the average access delay is also increasing, to a certain intensity, higher than the case in priority-based algorithm. Meanwhile, according to the priority from low to high, their access time delays are significantly reduced, and access time delay of high priority users is

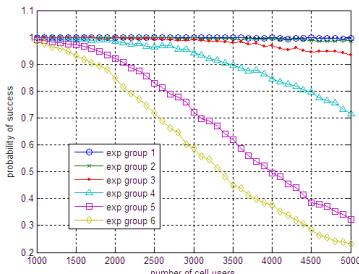
significantly lower than that of the low priority users. When users' capability is large, but the access delay is also obvious lower than that in general algorithm.

We introduce the priority distinguish. Users will be divided into several different priority levels, and the different priority users access the corresponding different sub-timeslot frames. As above, this paper uses three different functions to divide the priority. The distribution of the different priority users numbers in each algorithm as shown in Fig.5.

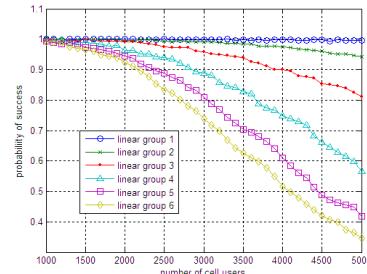


**Fig. 5.** The subscribers of priority group in three priority algorithms

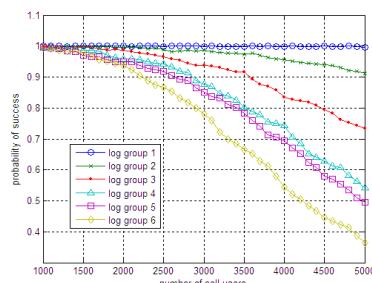
As shown in Fig.5, “linear” is the distribution of each level of the priority user in linear function algorithm, “exp” is the distribution of each level of the priority user in exponential function algorithm, “log” is the distribution of each level of the priority user in logarithmic function algorithm.



**Fig. 6.1.** Exponential function algorithm



**Fig. 6.2.** Linear function algorithm



**Fig. 6.3.** Logarithmic function algorithm

**Fig. 6.** Comparison of access success rate in three priority algorithms

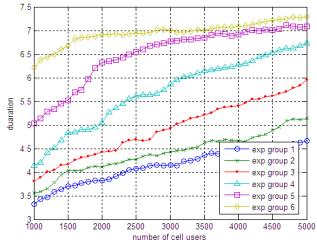
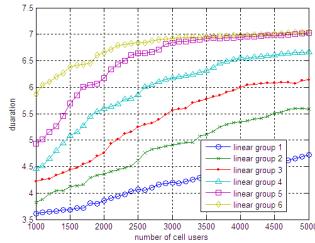
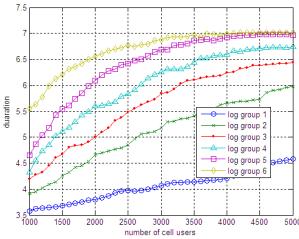
**Fig. 7.1.** Exponential function algorithm**Fig.7.2.** Linear function algorithm**Fig.7.3.** Logarithmic function algorithm**Fig. 7.** Comparison of time delay of access in three priority algorithms

Fig.6 and Fig.7 show the comparison of the success rate and time delay of the users' accession in the simulation environment, respectively based on the three priority algorithm. “exp (linear or log) group  $i (i=1, \dots, 6)$ ” are the different priority user in the priority algorithm, “priority group 1” is the user whose priority is highest.

Based on Fig.6, we comparing the access success rate of the three algorithms, the most obvious effect is exponential function algorithm, linear function algorithm is moderate, logarithmic function is the most moderate algorithm. The access success rate of high priority users in exponential function algorithm is higher than the other two algorithms. The access success rate of low priority users is relatively lower. So exponential function algorithm is more suitable to the requirement of high priority group, it provide a higher QoS support for the high-priority users. Even if confronting with the condition which access capacity is great, high-priority users also can smoothly access. Also we can choose logarithmic function algorithm when all users need a better QoS in the trunking group.

Fig.7 shows, comparing the access delay of the three algorithms, the most obvious effect is exponential function algorithm, the better is linear function algorithm, and then logarithmic function algorithm. Compared with the other two algorithms, the access delay of high priority users in the exponential algorithm is shorter, and the access time delay of low priority users is relatively longer. So the exponential algorithm is the best suitable for the high requirements in high priority groups.

## 5 Conclusion

In this paper, the random access algorithm based on priority is presented, which introduced the priority, the high priority user is priority to occupy the uplink traffic

channel, improving the user's access success rate, optimizing the QoS of network overall, meeting the requirements of high priority user have a higher success rate. Based on these problems, three kinds of algorithms were given to divide the priority in this article, and compared the merits of three algorithms by simulation experiments, in which exponential algorithm is the best suitable for high requirements of priority in the trunking system.

**Acknowledgments.** This work was supported by National Science & Technology Major Program (2012ZX03004009), the special fund of Chongqing key laboratory (CSTC) and the transformation project of excellent achievement of Chongqing Municipal Education Commission (Kjzh11206).

## References

1. Xu, X.-T.: The principles and applications of digital mobile trunking communication system. The People's Posts and Telecommunications Press (2008)
2. Xie, X.-Z.: TD-SCDMA the third generation mobile communications system technology and implementation. Electronic Industry Press (2004)
3. Luo, Y.-J., Duan, H.-G.: The analysis of random access process in TD-SCDMA system. Guangdong Communication Technology (January 2005)
4. Zhou, H.-J., Yan, X.-L., Xiong, S.-M., Xie, X.-Z.: RACH Performance Analysis in TD-SCDMA system. Journal of Chongchong University of Posts and Telecommunications 14(4), 19–23, 64 (2002)
5. 3GPP TS 25.331 Technical Specification 3rd Generation Partnership Project. Technical Specification Group Radio Access Network. Radio Resource Control (RRC). Protocol Specification
6. Zheng, L., He, J.-H., Yang, Z.-K., Cheng, W.-Q.: Performance Evaluation of IEEE 802.11 Wireless LAN Supporting Multi-Priority Services. Computer Science 31, 46–49 (2004)

# Research on Improved DV-HOP Localization Algorithm Based on the Ratio of Distances

Yan Hu<sup>1</sup>, Zhilong Shan<sup>1</sup>, and Hua Yu<sup>2</sup>

<sup>1</sup> School of Computer, South China Normal University, Guangzhou 510631, China  
sunnysz1@163.com

<sup>2</sup> School of Electronic and Information Engineering  
South China University of Technology Guangzhou 510640, China  
yuhua@scut.edu.cn

**Abstract.** In this paper, we study the Range-free localization algorithm using the ratio of distances based on the DV-HOP algorithm in wireless sensor networks. An improved SDDV-HOP (Shortest Distances DV-HOP) algorithm is proposed by modifying the network average hop distance. The simulation results show that SDDV-HOP algorithm can improve the localization accuracy greatly compared with DV-HOP algorithm.

**Keywords:** Wireless sensor network, The Shortest-path, DV-HOP, Localization.

## 1 Introduction

For most wireless sensor network applications, the data is meaningless if we can't know the location of sensors which obtained it. Determining the location where the incident happened or the location of node which got the information is one of most basic functions. What's more, the localization accuracy is related with the wireless sensor network applications effectiveness. Due to cost constraints, it is impossible every node in the wireless sensor network equipped with GPS. Generally only a few nodes possess precise localization capability using GPS and the remaining nodes estimate their own location through specific localization algorithm. Therefore, study on node localization algorithm for wireless sensor network has important theoretical significance and practical value [1].

According to whether there is a need for actual measurement of the distance between nodes or nodes' relative angle, self-localization algorithm for wireless sensor network can be divided into the Range-Based localization algorithm and Range-Free localization algorithm [2].

Range-Based localization algorithm determines the location of unknown node by measuring the actual distance between adjacent nodes or actual location of the nodes. Range-Free localization algorithm mainly relies on network connectivity without measuring the distance between adjacent nodes or locations of the nodes actually. Thus, the hardware requirements are reduced and the nodes cost is more suitable for large-scale wireless sensor networks. Although the Range-Free localization algorithm

lacks of accuracy, it can meet the most requirements of wireless sensor network applications and attract more and more attention [3]. The DV-HOP algorithm is one of the typical representatives of Range-Free localization algorithm [4]. According to the disadvantage of DV-HOP algorithm and the existing DV-HOP modified algorithm, this paper proposed the weighted shortest-path distance and straight distance ratio between nodes to modify the average hop distance and conducted simulation.

The paper is organized as follows. Section 2 describes the DV-HOP algorithm. In Section 3, the proposed SDDV-HOP algorithm is presented. The performance of the proposed method is evaluated in Section 4. We conclude our paper in Section 5.

## 2 DV-Hop Algorithm

In this section, we introduce the DV-HOP scheme firstly which consists of three stages. In the following stages, the sensor nodes without location information are referred as unknown nodes [5].

### 2.1 Get the Minimal Hopcount between Anchor Nodes and Unknown Nodes

Each anchor node broadcasts a beacon message throughout the network containing its location information with a hopcount initialized to zero. The nodes having received the information will record the minimal hopcount to every anchor node and forward it to the neighbor node, ignoring the larger hopcount at the same time.

### 2.2 Calculation the Actual Distances between Unknown Nodes and Anchor Nodes

Every anchor node estimates average hopsizes ( $Hopsize_i$ ) according to the recorded hopcount and position information to other anchor nodes.  $Hopsize_i$  is defined as

$$Hopsize_i = \frac{\sum_{i \neq j} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{\sum_{i \neq j} h_{ij}} \quad (1)$$

where  $(x_i, y_i)$  and  $(x_j, y_j)$  are the coordinates of anchor nodes  $i$  and  $j$ .  $h_{ij}$  is the hopcount between anchor node  $i$  and all anchors nodes. Anchor nodes broadcast this information to the network by controlled flooding. The unknown nodes record the first received average hopsizes, then forward it to the neighbors, and drop the subsequent ones simultaneously. Finally, the unknown nodes compute the distances to the anchor nodes through multiplying the received average hopsizes by the hopcount. The distance  $d_{ij}$  can be expressed as

$$d_{ij} = Hopsize_i * h_{ij} \quad (2)$$

### 2.3 Compute the Coordinates of Unknown Nodes

The unknown nodes ( $w$ ) then have estimated distances to anchor nodes, in meters, which can be used to perform the tri-lateration method. Suppose a node  $u$  located at coordinates  $(x_w, y_w)$  utilizes  $n$  anchor nodes to estimate its position. Let  $(x_i, y_i)$  be the coordinates of the  $i^{\text{th}}$  anchor node. The estimated distance  $d_{wi}$  can be related with the coordinates  $(x_w, y_w)$  and  $(x_i, y_i)$  by (3).

$$d_{wi} = (x_w - x_i)^2 + (y_w - y_i)^2, i = 1, 2, \dots, n \quad (3)$$

The coordinate of the node  $u$  could be estimated with the Least-Square solution defined as follow

$$X = (A^T A)^{-1} A^T b \quad (4)$$

where

$$A = -2 * \begin{bmatrix} x_1 - x_n & y_1 - y_n \\ x_2 - x_n & y_2 - y_n \\ \vdots & \vdots \\ x_{n-1} - x_n & y_{n-1} - y_n \end{bmatrix} \quad (5)$$

$$b = \begin{bmatrix} d_1^2 - d_n^2 - x_1^2 + x_n^2 - y_1^2 + y_n^2 \\ d_2^2 - d_n^2 - x_2^2 + x_n^2 - y_2^2 + y_n^2 \\ \vdots \\ d_{n-1}^2 - d_n^2 - x_{n-1}^2 + x_n^2 - y_{n-1}^2 + y_n^2 \end{bmatrix}$$

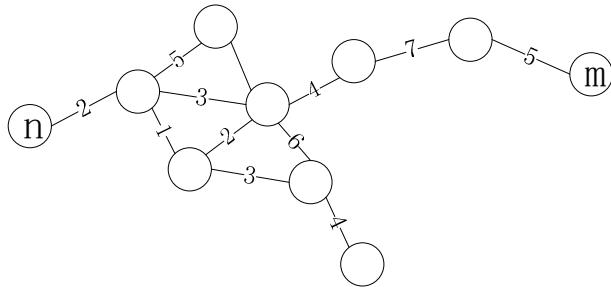
## 3 SDDV-HOP Algorithm

### 3.1 All-Pairs Shortest-Path Principle

If you have more than one surname, please make sure that the Volume Editor knows how you are to be listed in the author index.

In this paper, the idea originates from the ratio of shortest-path distance and straight distance between nodes. In graph theory, the shortest path problem is the problem of finding a path between two nodes in a graph so that the sum of the weights of its constituent edges is minimized [6].

There are several variations according to whether the given graph is undirected, directed, or mixed. For undirected graphs, the shortest path problem can be formally defined as follows. Given a weighted graph as Fig. 1(that is, a set  $N$  of nodes, a set  $E$  of edges, and a real-valued weight function  $f : E \rightarrow \mathbb{R}$ ), and elements  $n$  and  $m$  of  $N$ , find a path  $P$  from  $n$  to a  $m$  of  $N$  so that  $\sum_{p \in P} f(p)$  is minimal among all paths connecting  $n$  to  $m$ .



**Fig. 1.** A weighted graph

Sometimes, the problem is also called the single-pair shortest path problem to distinguish it from the all-pairs shortest path problem. The single-pair problem is always solved by the Dijkstra's algorithm. The all-pairs shortest path principle is actually the approach of running a single-pair shortest path algorithm on all relevant pairs of nodes. However, in order to reduce the algorithm complexity, we adopt the Johnson's algorithm to build the shortest path among all nodes. In this paper, the weights of its constituent edges are the distances obtained from the RSSI-distance model between nodes.

### 3.2 RSSI-Distance Model

Based on the current study, there have been three RSSI model: the optimal model, linear model and theoretical model. Under the existing experimental environment and the requirements of the positioning accuracy, we choose the linear model. RSSI linear model uses a pair of parameters  $\alpha$  and  $\beta$  to estimate the distance between nodes, as shown in (6), where  $r$  is the value of RSSI Ranging [7].

$$\hat{d} = \alpha * r + \beta \quad (6)$$

### 3.3 Shortest Path Distance Applied in the DV-HOP Algorithm

To improve the location accuracy of the DV-HOP algorithm [1, 8-10], we propose a SDDV-HOP (Shortest Distances DV-HOP) algorithm by modifying the network average hop distance.

In the primitive algorithm, the average hopsize equals straight-line distance divided by minimum hopcount between anchor nodes. Actually, the straight-line distance is quite different from the sum of distances between the nodes on the smallest hopcount path. The hopsize obtained from (1) can not get the most suitable hopsize to the network nature.

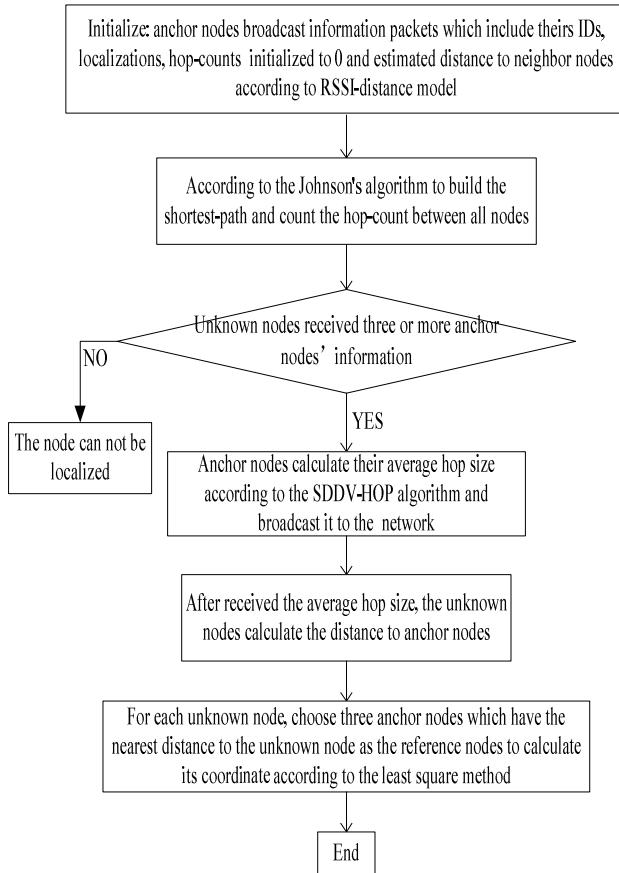
In order to solve this problem, we choose the shortest-path distance rather than the smallest hopcount path, and view the ratio between shortest-path distance and the straight-line distance as the contribution degree of straight-line distance to modify the whole network average hopsize when getting it. Hence,  $Hopsize_i$  can be defined as

$$Hopsize_i = \frac{\sum_{i \neq j} D_{ij} * \frac{1}{w_{ij}}}{\sum_{i \neq j} h_{ij} * \frac{1}{w_{ij}}} . \quad (7)$$

where  $S_{ij}$  is the shortest-path distance between the reference nodes  $i$  and  $j$ .

As in Fig. 1, when we build the shortest path, we use the estimated distance obtained from RSSI-distance model as the weights of the edges.

The process of the improved algorithm is shown in Fig. 2.



**Fig. 2.** SDDV-HOP algorithm flow chart

## 4 Simulation Results

In this section, we study the new localization scheme by computer simulation. The network is a square with side length of 100m as shown in Fig.3. Let the node distribution be distributed randomly with 20m transmission range in the network.

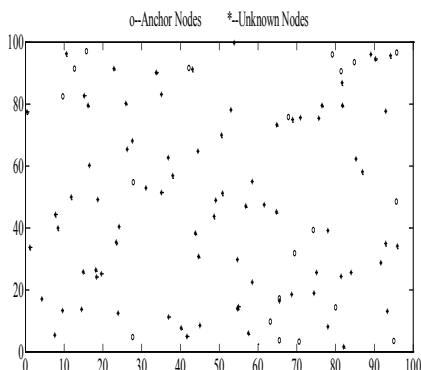
We use matlab as simulator to implement the network scenario and determine the location results. To measure the accuracy of location, the error  $E$  is defined as

$$E = \frac{\sum_{i=1}^{UNAmount} \sqrt{(\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2}}{UNAmount} / R \quad (9)$$

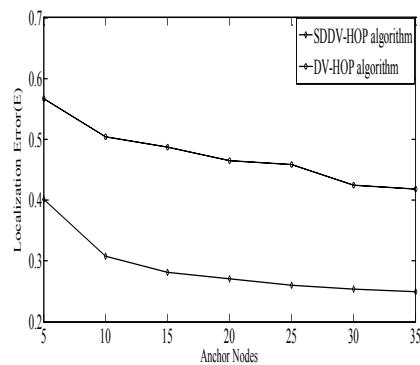
where UNAmount is the number of the unknown nodes in the network.  $R$  is the communication radius of nodes.

The communication radius of each node is assumed to be 20m. The anchor nodes are selected from all of the nodes proportionally. In the figures, the localization error in the y-axis is the average localization error  $E$  after running the algorithm for 20 times.

As shown in Fig. 4, the SDDV-HOP algorithm achieves better performance than the DV-HOP algorithm. For instance, at 20% of anchor nodes, SDDV-HOP algorithm exhibits an average location error of about 28%, which are 22% more accurate than the DV-HOP algorithm. As the anchor nodes increase in number, the average hopsize conforms better to the network nature. So as expected, the location error decreases with the increasing of percentage of anchor nodes.



**Fig. 3.** Node Distribution



**Fig. 4.** Localization error vs Anchor Nodes

Fig. 5 shows the performance of DV-HOP scheme by the effects of nodes density on location accuracy. The nodes density is the average numbers of nodes in the circular region in given radius. In this simulation, the node density varies from 50 to 300 sensor nodes with an interval of 50, and we fix the number of anchor nodes to 15, with a radius equal to 20m. The results show the error decreases with the increasing of the node density. Especially, the location error reaches to the stage of stability as the number keeps increasing when the number of the sensor nodes reaches a fixed value. When the number of sensor nodes is much more, SDDV-HOP algorithm is 15%~25% more accurate than the DV-HOP algorithm. When the number is small, SDDV-HOP algorithm is still 4%~10% more accurate than the DV-HOP algorithm.

Fig. 6 shows the performance of SDDV-HOP algorithm by the effects of communication radius and the anchor nodes. As the result shows, the location error decreases as the anchor nodes and communication radius increasing. When the anchor nodes are fixed, we increase the communication radius, and all the sensor nodes have more neighbors to be used for location. Therefore, the location error decreases as the communication radius increases.

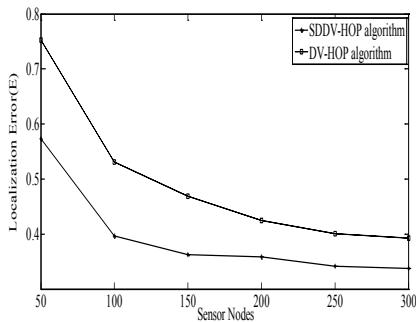


Fig. 5. Localization error vs Sensor nodes

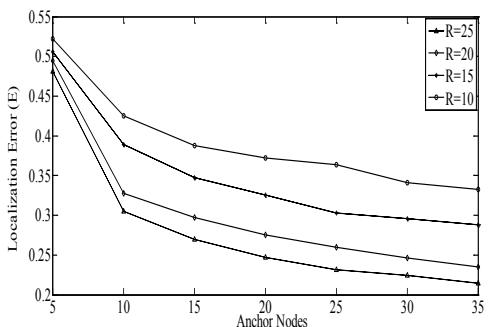


Fig. 6. Localization error vs communication radius

## 5 Conclusions

A SDDV-HOP algorithm for wireless sensor networks is proposed in this paper. In this algorithm, the shortest-path principle is introduced. The ratio of the shortest-path distance and straight-line distance is used as the weights to modify average hop size in the network. The simulation results show that SDDV-HOP algorithm can improve the location accuracy compared with the DV-HOP algorithm.

**Acknowledgment.** This work was supported by National Natural Science Foundation of China (No.61071212), and Science and Technology Planning Project of Guangzhou (No. 11C42090538), and Open Project Program of Key Laboratory of wireless communication network and terminal of Guangdong Higher Education Institutes (No. KLB08002), and Natural Science Foundation of Guangdong Province (No. 10451063101006313).

## References

1. Zhang, J.-G., Li, W., Cui, D., Sun, X., Zhou, F.: Study on Improved DV-Hop Node Localization Algorithm in Wireless Sensor Network. In: 2010 the 5th IEEE Conference on Industrial Electronics and Applications (ICIEA), vol. 6, pp. 1855–1858 (2010)
2. Niculescu, D., Nath, B.: Ad-hoc positioning system (APS). In: Proc. of the IEEE GLOBECOM, San Antonio, pp. 2926–2931 (2001)
3. Li, C.-R.: The Self-Location Technology Research for Wireless Sensor Network. Southwest Jiaotong University, Master's Paper (June 2006)

4. Nasipuri, A., Li, K.: A directionality based location discovery scheme for wireless sensor networks. In: Proceedings of ACM International Workshop on Wireless Sensor Networks and Applications, pp. 105–111. Association for Computing Machinery, New York (2002)
5. Brito, L.A., Liu, Y., Garcial, Y.: An Improved Error Localization on DV-Hop Scheme for Wireless Sensors Networks. In: 2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE), vol. 2, pp. 80–84 (2010)
6. [http://en.wikipedia.org/wiki/Dijkstra's\\_algorithm](http://en.wikipedia.org/wiki/Dijkstra's_algorithm)
7. Yi, T.-T., Fang, Z., Li, R.-X.: RMADV-Hop: An Improved DV-Hop Localization Algorithm. In: 2010 Seventh International Conference on Information Technology, vol. 6, pp. 939–943 (2010)
8. Niculescu, D., Nath, B.: DV-based positioning in ad hoc network. *Telecommunication Systems* 22(1), 267–280 (2003)
9. Zhang, J., Wu, Y.-H., Shi, F., Geng, F.: Localization algorithm based on DV-HOP for wireless sensor networks. *Journal of Computer Application* 30(2), 323–326 (2010)
10. Lin, J.-Z., Liu, H.-B., Li, G.-J., Liu, Z.-J.: Study for improved DV-Hop localization algorithm in WSN. *Application Research of Computers* 26(4), 1272–1275 (2009)

# **Survivability Evaluation of Cluster-Based Wireless Sensor Network under DoS Attack**

Chunjie Chang, Changhua Zhu, Honggang Wang, and Changxing Pei

State Key Lab of Integrated Services Networks, Xidian University,  
Xi'an, 710071 China

chanchunj@126.com, {chhzhu, chxpei}@xidian.edu.cn,  
whg\_xian\_cn@163.com

**Abstract.** Survivability of Wireless Sensor Network(WSN) can be defined as the capability to fulfill its mission in the presence of attacks, accidents and failures. In this paper, we propose a survivability model for cluster-based WSN, in which the state of each cluster is regarded as a stochastic process based on a semi-Markov process (SMP) and Discrete Time Markov Chain (DTMC). The isolation problem between clusters is discussed. Quantitative survivability is obtained based on  $k$  connectivity. Numerical results show our model is effective.

**Keywords:** semi-Markov process, Discrete Time Markov Chain, Survivability.

## **1 Introduction**

WSN is vulnerable and prone to failure because of its characteristics such as wireless transmission and unmanned on duty. Network survivability of WSN is recently drawing ever-increasing attention. The suitable survivability model should be proposed. In the literature, the survivability of WSN under DoS attack was analyzed and the degree of basic services under DoS attack between cluster head nodes was evaluated [1]. A single cluster was discussed by using the semi-Markov and the Discrete Time Markov Chain [2]. The survivability of wireless Ad Hoc networks with node misbehaviors and failures had been evaluated which was proposed [3,4]. In this paper our model is built on Kim's model [2].

The remainder of this paper is organized as follows: In Section 2 the model of WSN is built. In Section 3 the isolation problem between cluster head nodes is discussed. In Section 4 the formula of survivability is obtained. Simulation results are provided In Section 5. Section 6 concludes this paper.

## **2 Model of Wireless Sensor Network**

### **2.1 WSN Topology**

Cluster-based WSN have three types of nodes: sink node, cluster head node and common sensor node. In our analysis, we consider a hierarchical architecture with

dynamic topology rather than a flat one, in which one or more clusters connected with the sink node and the sink node is connected to the legacy network. For a cluster based WSN, we prefer to suppose that the cluster head node acts as an agency which is responsible for the whole inner-cluster data transmission between the common node and the sink node. For this reason, in our analysis, we think the behavior of the cluster head node is equivalent to that of all the common sensor nodes.

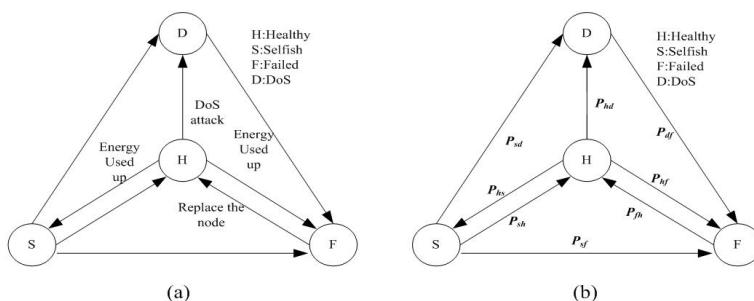
## 2.2 Semi Markov Process for Cluster Head Node

In WSN, the state of a cluster head node exhibits random behavior, since the transition of a state can be influenced by various reasons. The attack causes the sojourn time of some states to be non-exponential. Simple Markov chain can't represent a cluster head node. So a cluster head node is modeled as semi-Markov process.

In this work, we assume that all cluster head nodes operate independently in the following four states [5].

- Healthy state(H). In this state, the cluster head node can communicate well with common sensor nodes in its own cluster and can translate the packets which are sent by other cluster head nodes.
- Selfish state(S). In this state, the cluster head node can only communicate with common sensor nodes in its own cluster for the sake of saving energy.
- DoS state(D). In this state, the cluster head node which is compromised by DoS attack(here we consider the Blackhole attack) responses to other cluster head nodes a fake RREP message immediately claiming that it is in the optimal path or only one-hop away to the destination node. But the Dos cluster head node will dump all packets without proper countermeasures.
- Failed state(F). In this state, the energy of the cluster head node used up, the cluster head node is dead and can't communicate with other cluster head nodes and common sensor nodes.

Based on the description above, we define a state space  $S=\{H, S, D, F\}$ and model the process of behavior transitions by semi-Markov chain. Fig.1(a) depicts SMP model for a cluster head node.



**Fig. 1.** Embedded DTMC of SMP model Semi and Markov Process for a cluster head node

### 2.3 Discrete Time Markov Chain of SMP

The transition from one state to another in an SMP can be thought as two steps transition logically[7]. In the first stage, the process remains in state  $i$  for an amount of time given by  $T_i(t)$ , where  $T_i(t)$  is the sojourn time distribution of state  $i$ . In the next stage, the process changes its state from  $i$  to  $j$  with the transition probability  $p_{ij}$ .

The state  $X_n$  is denoted at transition time  $t_n$ , then

$$P_r(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) = P_r(X_{n+1} = x_{n+1} | X_n = x_n) \quad (1)$$

where  $x_i \in S$   $0 \leq i \leq n + 1$ . Therefore, an SMP{Z(t),  $t \geq 0$ } is used to model cluster head node behavior transitions, which is defined by

$$Z(t) = X_n, \forall t_n \leq t \leq t_{n+1} \quad (2)$$

In (2),  $Z(t)$  refers to the state of the process during the period from the most recent transition,  $\{X_n\}$  is called the embedded Markov chain of the process SMP{Z(t),  $t \geq 0$ }.

The state probability vectors of the DTMC  $\{X_n\}$  and SMP  $Z(t)$  are  $\pi$  [ $\pi_h$ ,  $\pi_s$ ,  $\pi_d$ ,  $\pi_f$ ] and  $p$  [ $p_h$ ,  $p_s$ ,  $p_d$ ,  $p_f$ ]. Fig.1 (b) depicts the embedded Discrete Time Markov Chain.

$P$  is the transient matrix of DTMC expressed as

$$P = \begin{bmatrix} 0 & p_{hs} & p_{hd} & p_{hf} \\ p_{sh} & 0 & p_{sd} & p_{sf} \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

where  $p_{ii}=0$  means that  $\{X_n\}$  only has transitions from a state to another state. A transition probability of zero like  $p_{ds}=0$ ,  $p_{dh}=0$  means a DoS cluster head node will not become a healthy or a selfish.

To obtain the state probabilities, we need to solve the equation.

$$\bar{\pi} = \bar{\pi}P, \sum_{i \in S} \pi_i = 1 \quad (3)$$

The steady state probabilities of SMP is computed by using equation

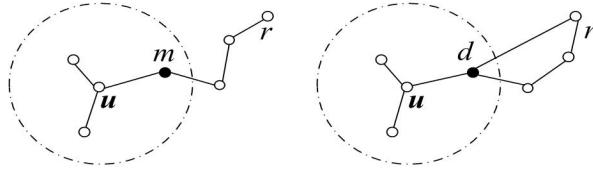
$$P_j = \frac{\pi_j E[T_j]}{\sum_{j \in S} \pi_j E[T_j]} \quad (4)$$

Where  $T_i$  means sojourn time in  $i$  state and  $E[T_i]$  means mean sojourn time in  $i$  state.

## 3 Cluster Head Node Isolation

A trivial case of cluster head node isolation occurs when a cluster head node becomes failure. We study more general cases of node isolation by considering the following three scenarios:

- (1) The effect of failed neighbors. In Fig.2(a) suppose a cluster head node  $m$  is a failed node. When node  $u$  initiates a route discovery to another node  $r$ , the failed neighbor  $m$  is unable to communicate with other cluster head node and common sensor node in its own cluster. In this case,  $u$  is isolated by its failed neighbors[4].
- (2) The effect of selfish neighbors. Note that selfish cluster head nodes can still communicate with common nodes in its own, which is different from failed nodes.



**Fig. 2.** (a) Failed and Selfish neighbors      (b) DoS neighbors

- (3) The effect of DoS neighbors. As illustrated in Fig.2(b), suppose AODV is used as routing protocol. When cluster head node  $u$  discovers the route to cluster head node  $r$  by broadcasting RREQ messages, a Blackhole neighbor, say,  $d$ , can respond to  $u$  with a fake RREP message immediately claiming that it is in the optimal path to  $r$ . Consequently,  $u$  selects  $d$  as the next hop and sends data to it, but  $d$  will just dump all packets without proper countermeasures [6].

Let  $n_s(u), n_f(u), n_d(u)$  are the numbers of  $u$ 's selfish, failed and DoS neighbors. The degree  $D$  is defined as the number of neighbors. The connectivity degree  $D_c$  is defined that the cluster head node has  $k$  healthy node and do not have the DoS neighbors.

**Proposition 1.** Given cluster head node  $u$  with neighbors  $d$ ,  $n_s(u)+n_f(u)=d$  or  $n_d(u)>=1$ , the connectivity degree  $D_c$  is zero. The probability with which  $D_c$  is zero is

$$\begin{aligned} P_r(D_c=0|D=d) &= P_r(n_d \geq 1|D=d) + P_r(n_s + n_f = d|D=d) \\ &= 1 - (1 - P_B)^d + (1 - P_h - P_B)^d \end{aligned} \quad (5)$$

Where,  $P_B$  is the probability of launching Dos attack and  $P_h$  is the probability of a cluster head node being in a healthy state.

## 4 Network Survivability

The network survivability of WSN is defined as the probability of all cluster head nodes except failed cluster head nodes are  $k$  connected to the network. For a WSN including  $M$  cluster head nodes [3].

$$NS_k(M) \approx P_r \left( \bigcap_{u \in M_a} D_c(u) \geq k \right) \quad (6)$$

Where  $M_a$  is the subnetwork of  $M$  induced by cluster head nodes expect failed cluster head nodes.

$$P_r(D_c=k|D=d) = P_r(n_h=k, n_b=0|D=d) + P_r(n_s+n_f=d-k|D=d) = \binom{d}{k} P_h^k (1-P_h-P_B)^{d-k} \quad k \geq 1 \quad (7)$$

In order to get the network survivability we need to solve  $P_r(D_c \leq k)$

$$P_r(D_c \leq k) = \sum_{d=0}^{\infty} P_r(D_c < k | D = d) P_r(D = d) \quad (8)$$

Now we need to find  $P_r(D=d)$ , cluster head nodes are associated with a poisson process  $u_a=\rho\pi r^2$  with density  $\rho=N(1-P_f)/A$ ,  $r$  is the transmission range of cluster head nodes,  $A$  is the simulation area.

$$P_r(D = d) = \frac{(u_a)^d}{d!} e^{-u_a} \quad (9)$$

According (5)-(8), we get the survivability formula is

$$\begin{aligned} NS_k(M) = & \left[ \frac{\Gamma(k, u_a)}{\Gamma(k)} + e^{-u_a} P_B \left( 1 - \frac{\Gamma(k, u_a (1 - P_B))}{\Gamma(k)} \right) \right. \\ & \left. - e^{-u_a} P_B \frac{\Gamma(k, u_a P_h)}{\Gamma(k)} \right]^{N(1 - P_f)} \end{aligned} \quad (10)$$

where  $\Gamma(x) = (x-1)!$ ,  $\Gamma(h, x) = (h-1)! e^{-x} \sum_{l=0}^{h-1} x^l / l!$ .

## 5 Simulation

In this work, we use NS2-v2.34 and matlab-v6.0 to perform the simulation. Firstly, the transient matrix of DTMC is given as following [3].

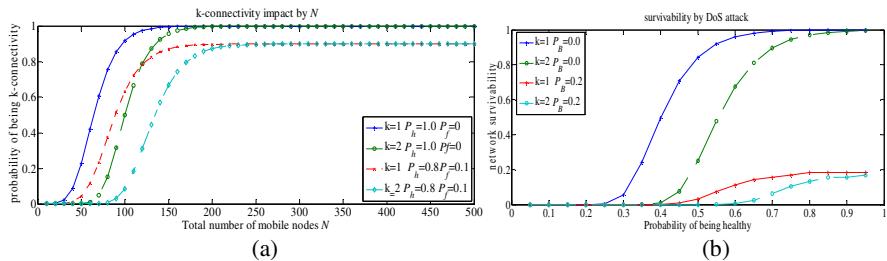
$$P = \begin{bmatrix} 0 & 0.525 & 0.071 & 0.404 \\ 0.756 & 0 & 0.022 & 0.222 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

The parameters of cluster-based WSN are shown in Table 1.

**Table 1.** Default Values of Parameter

<i>Simulation area</i>	500*1000
<i>MAC</i>	IEEE 802.11
<i>Transmission range</i>	100 m
<i>Cluster head node</i>	10-500
<i>Initial energy</i>	600Ws-2500Ws

The effect of the number of the cluster head nodes N is discussed, which is set from 0 to 200 and can confirm the coverage of the simulation area. For  $P_h=1$ ,  $P_f=0$  and N is 150 or larger, the survivability is 1, there is no isolation to the problem. The variations of the survivability with the number of N are shown in Fig.3 (a).



**Fig. 3.** The effect of different number of cluster head nodes and DoS attack on survivability

The effect of survivability of WSN under DoS attack is shown in Fig.3 (b). In experiment, we set the initial energy is 1000. The survivability rapidly drops to 0.18 when the probability of DoS attack is 0.2. The reason is that a Blackhole cluster head node can isolate all its neighbors.

## 6 Conclusions

In this paper, a survivability model of cluster based WSN is proposed. We begin from the behaviors of cluster head nodes by employing a semi-Markov process. Then the isolation problem of cluster head node is analyzed. We also discuss the connectivity degree of a single cluster head node. The survivability of WSN is obtained by the connectivity of cluster head nodes. Simulation results show that DoS attacks decrease the survivability quickly and the survivability can be improved by properly adding node energy and the number of cluster head nodes.

**Acknowledgement.** This work is supported by National Natural Science Foundation of China (No.61072067), The State Key Specialized Scientific Projects (2009ZX03007-003), Shanxi Key Industrial scientific and technological Project (2009K01-46), the 111 Project (B08038), the Fundamental Research Funds for the Central Universities (K50510010004).

## References

- [1] Jiang, Z.Q., Shu, Y., Wang, L.M.: Survivability Evaluation of Cluster-Based Wireless Sensor Network. In: 5th International Conference on Wireless Communications, Networking and Mobile Computing, pp. 1–4. IEEE press, New York (2009)
- [2] Kim, D.S., Shazzad, K.M., Park, J.S.: A framework of survivability model for wireless sensor network. In: Proc. 1st International Conf. Availability, Reliability Security (ARES), pp. 515–522. IEEE (2006)
- [3] Santi, P.: Topology Control in Wireless Ad Hoc and Sensor Networks. John Wiley & Sons (2006)
- [4] Xing, F., Wang, W.: Modeling and Analysis of Connectivity in Mobile Ad Hoc Networks with Misbehaving Nodes. In: Proc. IEEE Int'l Conf. Comm., pp. 1879–1884. IEEE press, New York (2006)
- [5] Xing, F., Wang, W.: On the Survivability of Wireless Ad Hoc Network with Node Misbehaviors and Failures Modeling. IEEE Transactions on Dependable and Secure Computing 7(3), 284–299 (2010)
- [6] Buttyan, L., Hubaux, J.-P.: Stimulating Cooperation in Self-Organizing Mobile Ad Hoc Networks. Mobile Networks and Applications 8(5), 579–592 (2003)
- [7] Iibriq, J., Mahgoub, I.: Cluster-Based Routing in Wireless Sensor Networks: Issues and Challenges. In: Proc. of the 2004 Sym. on Performance Evaluation of Computer Telecommunication Systems, pp. 759–766. The Society For Modeling and Simulation (2004)

# Research on Ubiquitous Network Technique and Application

Ping Zhu and Xiaofei Xu

School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China  
fendicmm@sina.com

**Abstract.** Under the increasing external world communicating with each other by the reason, concept, development of ubiquitous information network research on intelligence terminal system, foundation network technique, application technique etc. In this paper we sum up the characteristic, question and the direction of ubiquitous network and then analyze the case of its application. Finally we put forward the assumption to construct the ubiquitous network and the trend of the future developments.

**Keywords:** ubiquitous Computing, ubiquitous network, Ambient Intelligence, Internet of things, RFID, Sensor Network, M2M, Global Information Grid(GIG).

## 1 Introduction

Nowadays, the world is made to be apperceived and measurable through globalization, network and intelligence. The sensor has became the basic component in digital age before 60 years, each person have near upon one billion sensors until now, and the cost of each sensor is just one over ten million cents. The users of mobile phone are more than four billion; the users of network will quickly become two billion; the label of Radio Frequency Identification (RFID) becomes 30 billion in global world. More and more people will focus on the natural and manual systems to realize the real time perception and improve the development and advancement of society. In real world, there are intercommunion between systems and objects. The computation is abroad applied in every aspect in society besides the computer, any people, object, process or service in any scale organization are realized by the digital apperceive and network, such as high performance computer, cloud computation etc., which controls end-user equipment, sensor and connects them to a powerful and uniform background system. So the various data are transformed to information in time, and the information are further transformed into activities, these steps increase the efficiency of system, process, basic establishment, and productivity and response speed, at last make the system more brightness. Especially, it is urgent for the world to be perceived and intelligent in the nearby ten years, because there appear a series of problems, including the weather change, the frangibility of energy sources, environment, grocery, grocery supply chain, the safety (such as 9.11 Event),financial crisis and its legacy effect etc. The world will become smaller, flatter, more intercommunion, and wisdom. The

technology (such as Cable network, wireless network, sensor net, electron label, and wireless radio frequency etc.) makes the Information and Communication Technology (ICT) developed quickly and constructs the immanent networks[1].The paper analyses the development process of ubiquitous network, studies the primary techniques, emphasizes the application domains of ubiquitous network, and presents the development trend in future.

## 2 Development Course of Ubiquitous Network

### 2.1 The Motivation of Ubiquitous Network

In 1990, the computer scientist Mark Weiser investigate and research the usage pattern of computers, firstly presents that the computer pattern in future will be ubiquitous or immanence or universal, in another words, the users own the computation equipments in their life space, the equipments cooperate with each other and supply the services of computation and communication. The concept of Ubiquitous Computing is firstly presented [2].

Nomura Research Institute (NRI) Japan presents the Ubiquitous Network on the base of the former. After that, Japan and Korea start develop the ubiquitous network. Japan presents the next stratagem “u-Japan” under the condition of “E-Japan” has been finished in 2004. So Japan becomes one of the prior countries who describe the information stratagem using the “Ubiquitous” and construct the ubiquitous society. USA and Europe also start the research plan which is similar to ubiquitous network, such as Ambient Intelligence proposed by European Union, Pervasive Computing proposed by North America etc [3, 4].

### 2.2 The Concept of Ubiquitous Network

Ubiquitous network is used for apperceive objects and identification. Besides of sensor network, there are 2-dimensional bar code, 1-dimensional bar code and RFID [5, 6].

The intension of ubiquitous network is alternation between people and objects through sensor equipment and wireless network, it has the parts of internet and the internet of things at aspect of conformation, and also belongs to Intelligence system (reasoning, Situation modeling, Context processing, Business trigger) [7, 25]. Ubiquitous network realizes the information achievement, transformation, storage, perceive, decision-making and use on-demand and other services between people and people, people and object, object and object. M2M includes three types of communications: Machine-to-Machine, Man-to-Machine and Machine-To-Man. The machine is end-user equipment for information transforming independently through remote transmission, also includes data transformation, remote telemetry, remote sensing and remote control. M2M has been realized as commercial applications in USA, EU and Korea. Ubiquitous network has strong environment, contents, culture, language apperceive and intelligence, including telecom net, the internet and the next generation which combines various business, such as wired, wireless broadband access, sensor network and RFID etc.

Ubiquitous network can connect general high performance computer system, cloud computation and server, Workstations and microcomputer, and connect Mobile phone, PDA, Game Console, car navigation system, digital TV set, information appliance, RFID to network through IPv6 protocol. Ubiquitous network implement the integrative utilization of information, transact the text, data and static image, transfers dynamic image and voice, safety information exchange, swap and closed loop control, and meet the individuation requirement.

### 2.3 National and Overseas Development of Ubiquitous Network

Every country and state all present electron construction stratagem to improve the national power through ICT, such as E-Japan, e-Korea, E-Europe, I-Hub in Singapore, and u-Taiwan etc. These countries have understood that the ubiquitous information service can drive the development of many industries, including the ICT, aiming to increase the competition. They propose the stratagem programming, named by “Ubiquitous Network” [8-11].

International Telecommunications Union (ITU) presented the research problem about ubiquitous in 2008, it declared that they will efficiently and systematically advance the realization from the assumption and enlarge the scale from the local application from the view of standardization.

The software, hardware and the latest production of wireless mobile communications system are showed in Mobile World Congress (MWC2009, Feb.) at Barcelona. This congress images the phases and technique enlightened of ubiquitous network. IPTV proposed the P4P technology brought from the internet and telecom, which made the ubiquitous network and the internet combined together in March, 2009. With the advance of standardization, ubiquitous network is developing for depth. Europe intelligence system integration technology platform forecasts that there are four phases: the radio frequency is applied in logistics, retail and pharmacy before 2010, objects internet between 2010 and 2015, objects come into semi-intelligent between 2015-2020, objects come into intelligent after 2020. Although many technology of the internet of things is still development test phase, there is gap to the combination of heterogeneous systems, the connect of objects [12].

Ubiquitous network is very important for the economy development in China, drives development of information industry through construction of the network, accelerates the new product, industry, and comes into being new economic spin-offs. Ubiquitous network will enhance the service level of information and network, is used to solve the imbalance among regions, cities and countries. China has established the stratagem target of ubiquitous network, development project and implementation tactic. Our country organizes the Industry-Academia-Research (IAR) to strength the research of technology, production and standard. The typical representatives are U-Beijing, U-Qingdao. Our country pay more attention to support the integration of telecommunications networks, cable TV networks and the internet, the internet of things, and cloud computation. U-China will be integrated into information environment, and bring new opportunity for information industry with the target to the world information technology leading position.

### 3 Ubiquitous Network Study

#### 3.1 Research Contents

Ubiquitous network contains fastness/portable communication network, sensor network, broadcast TV network, the internet and the internet of things, which supply many general services. The primary problem is how to design the architecture of ubiquitous network. ITU also builds the ubiquitous network societies project and workshop besides traditional communication domain, in where, integration architecture is important study project. The research direction of ubiquitous network focuses on basic network, P2P, system architecture, equipment extension, mobility, context aware etc. [13, 25]. With the maturation of IP-based next generation, Ad Hoc and access network, the basic network is not research hotspot. P2P has been a bottleneck of ubiquitous network because of the finite address space of IPv4. IPv6 is used to solve the problem [14,15].

The architecture of Ubiquitous network includes: the technology and application object, system architecture; module and component; the interface of modules; data identifier (collection, transaction, transformation, storage, query etc.); service standard; information safety; personal privacy protection.

#### 3.2 Primary Technique

There are the key techniques in Ubiquitous network, they are: architecture, organization, content aware in the wireless environment, concomitance and cooperate in the isomeric wireless access network, mobility management in isomeric network, advance data management, Cross-realm cross layer optimization technique [16].

Wireless technique is integrated to support ubiquitous network, for example, Bluetooth, 3G/4G, Optical fiber application etc. ubiquitous network have all of types of networks, and isomeric network integrative environment. 3G/4G becomes the common platform of wireless communication using access instrument (such as WiMAX, WLAN) of local network. The other stone's throw access techniques like UWB, RFID has been integrated into ubiquitous networks. There are three types in the technique architecture: intelligent terminal system, basic network and application.

According to requirement and technique development, the standards about business requirement, next generation intelligent network architecture and its communication protocol and access technique are researched. The contracture of ubiquitous network depends on the existence and interaction of basic network, terminal cell, network application. The next generation network, sensor network, RF identification, object symbol are the new research contents. It's difficult for one department or enterprise to complete the ubiquitous network, because there are too many techniques and standards.

Aiming to accelerate the development of ubiquitous network, it's necessary for our country to construct a framework and gradually solve the problems: interaction, safety, resource, implementation approach, industry integration etc.

## 4 The Application of Ubiquitous Network

At present, the research about ubiquitous network are developed in USA, EU, Japan, Korea etc., the initially research directions contains bar code, RF and its applications on retail, logistics. But with the development of RF, sensor, short range communication and computer, the environment monitoring, Biological and medical treatment, intelligence equipment are added into research plan. For example, IMEC uses the GPS, RF to develop the environment monitor and advance industry monitor, they uses advance technology in medicine field to develop automated driving system and the microelectronics sensor for human body, which is remote control, small and low cost.

Cisco Systems, Inc. develops the project named “Smart Connected Buildings”, which saves 15% energy consume for NetApp. The smart earth from IBM have developed the application projects, which contain the smart power, smart medical treatment, smart traffic, smart bank, smart city etc. From the above, ubiquitous is applied into many aspects in society, especially in military affairs.

### 4.1 Application of Ubiquitous Network in the Social Activity

First task of ubiquitous network is application through constructing the basic infrastructure which connects the internet and people, people and computer, computer and computer, people and object, object and object. Every application environment based on the infrastructure contain a platform which combines 3G/4G and RFID to support electronic payment, logistics monitoring, traffic management, industry control, including the mobile phone, PDA, notebook computer. Sensor Network is composed of sensors, which are distributed and connected into wireless autonomy network [17, 18]. It can aware the environment parameters, such as temperature, shake, helps people to discovery and alarm the ecological disaster. NFC and RFID are raising technologies for short's throw connect. It supplies the functions, such as change for music, video, and address list; it also implements non-impact smart card or reader, mobile payment.

Ubiquitous network connects the terminal cell and Nervous system. These decrease the cost of energy sources and implement the smart management. It's important to some industry, including energy, government, medical treatment, retail, service, traffic etc.

Information service includes mobile information service, individual search, information recommend and filtrate in the ubiquitous network. Aiming to incorporate smart information management platform, it shares cloud storage service through existing network infrastructure, protocol, and wireless. By this way, the efficiency of information service will be enhanced.

### 4.2 Application of Ubiquitous Network in Military Affairs

Base on smart management and information infrastructure, the ubiquitous network is applied in military. Global Information Grid (GIG) is needed to meet the requirement of advantage of information and decision making in 《Joint Vision 2020》 . GIG is a short name for the operation concepts and contents in information matrix. The spirit of ubiquitous network is shown from the incipient application in military of the internet to

C2, C3, C3I, C4I, C4ISR, C4KISR, GIG [19-22]. The combat entities in centric warfare depend on the matrix. The autoimmunization information system focuses on the information resource about war based on its interoperability and integration.

JV2010 presents an advance battle space information system; information network sends the information from the sensors to army. The aware of battle space controls the grid and dictates judgment function; it integrates the direction, control, communication, computer, intelligence, monitoring, and reconnaissance together. Ubiquitous network creates a efficient environment, the users visits the data in anywhere and not need to information distribution. These data maybe come from many places, for example, the armament system from other military, data space, monitoring and secondary planet [23-24].

The ubiquitous network in military is a System of System (SoS), supplies the increment functions for global world. It realizes information transaction, storage and transfer; finishes the internet management, distribution, assurance etc. The architecture of ubiquitous network builds on the communication and computation system makes the different architecture information systems integrate to a single high usable system. The system can solve many problems likes encapsulation, integration and interoperability. It helps to collect, transaction, protect data and implement the communication among the entities in battles. In the Joint battle, IT infrastructure is needed to further satisfy the requirement of combat troops and support persons in global world. It has some capabilities, such as safety network environment, infrastructure which can be used in anywhere, and software sharing the data and applications [24]. The terminal cell of military ubiquitous network is composed of multiplicity, access methods and function cells. It helps to change the military. The change depends on gains of information account, information transaction, efficiency, cooperation, integration [22].

Therefore, military ubiquitous network is an information system consisting of the interoperate process part and communication part, which can provide better capability from all combat positions, and function as communication port to connect between the joint forces, allied forces, non-national defense users and the system, thus being able to achieve the integrated command and control for the joint force with advantages on information and strategy, improve the interoperate capability of the system, maximally optimize the broadband capability, and greatly improve the fighting ability of the joint force within the overall combat. This help guarantee conducting the combat by use of less people and more rapid speed. Under this situation, the dynamic planning and re-direction of the resources are conducted in standard method; It helps for Joint commanders, assigned forces and National Command Authority (NCA) to transfer the information about target, forces maneuvering, Equipment conditions, Supply standard and resource configuration in ordain time. It makes military have an aware ability to discovery, lockout, tracking and aim any motorial objects. The ability supports to long distance, High precision weapons, Precision strikes.

Combat troops can get usable information from distribute resources, and manage the support information. It collects, transact, storage, distribution, display the information among the organizations in the battle space. Because information transfer and exchange seamless in global world, it cooperate the task and execution from the distributive and different levels (such as stratagem, battle, tactics and affair). It supports the real time, connectedness, usability of information for decision. It supports

integrative, strong viability and permanent communication information for administrators, synthesis tactics alarm and army.

Under the uniform framework, military ubiquitous network realizes the integration of data net, sound net, video net, sensor net and campaign platform. According to connection among land base, marine base, air force base, space base, internet base together, the command can build the secrecy and interconnected network from planet to sensor, from solo soldier to command center. It supplies necessary data, application software for Communication ability for various Combat troops. It makes the sensor, information platform, weapon system construct an organic military system. By this way, it improves the Speed of Command, quicken the process of campaign, increase execution, integrates different various types of armies and enhance the capability of campaign.

## 5 Conclusion

Hu Jintao gives an important presentation at Academician congress in June.7, 2010. He points that “when we develop information science technology, the internet of things should be developed quickly, construct new internet, change the situation which is control by persons. We should share information, keep the information safety. The intelligent broad wireless net, advance sensor and display technology need to be developed. We need to construct the infrastructure of the sensor net, communication equipment, network HPC computation, intelligent software. According to credibility, low cost, the ubiquitous information network architecture is needed to construct”. Ubiquitous network need the induction of government, enterprise, investigation. On the technology, we not only append on the existing network, but also develop wireless network. There are many technologies, such as Wi-Fi, 3G, ADSL, FTTH, electron label, RF, these technology are actively developed and applied.

Ubiquitous network comes through IOT phase, coordinated phase, integration phase, implements connection, integration and extinction. From the view of technology, a lot of standards are proposed, the primary techniques has been on the right road. From the view of application, ubiquitous network has been attached importance by more and more countries. Concrete service has transformed from the life activity to enhance economy running efficiency, its features helps to move to new business pattern. It will enhance the informatization and applications of the results in the social life.

**Acknowledgments.** This work is supported by State 863 High-Tech Program (No. 2009AA01A402) of China. The author would like to thank the people's helpful suggestions for revising this paper.

## References

1. ITU Ubiquitous Network Societies,  
<http://www.itu.int/osg/spu/ni/ubiquitous/>
2. Weiser, M.: The computer for the 21st century. Scientific American 265(3), 94–104 (1991)

3. IST. The Ambient Networks (DB/OL) (2006),  
<http://www.ambientnetworks.org/>
4. Wu, Z.-H., Pan, G.: Pervasive Computing. The Development Report of China Computer Science Technology 2005, pp. 175–187. Press of Tushua University, Beijing (2006) (in Chinese)
5. International Telecommunication Union UIT. ITU internet reports 2005:the internet of things (2005)
6. Lesser, V.: Distributed sensor networks: A multiagent perspective. Kluwer Academic Publishers, Boston (2003)
7. Katarzyna, K., Mauricio, T., et al.: Sky computing. IEEE Internet Computing (2009)
8. Rhee, C.-W.: u-Korea, <http://www.obi.giti.waseda.ac.jp/ITU/2004/documents/PS4-2.pdf>
9. Lie, E.: Ubiquitous Network Societies—The Case of Singapore. In: ITU Workshop on Ubiquitous, Geneva Switzerland (April 2005)
10. Umino, A.: Japan's New IT Reform Strategy and u-Japan,  
[http://www.soumu.go.jp/main\\_sosiki/joho\\_tsusin/eng/presentation/pdf/071122\\_1.pdf](http://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/presentation/pdf/071122_1.pdf)
11. u-Taiwan, <http://www.utaiwan.nat.gov.tw/>
12. ITU. The Case Of Japan (EB/OL), <http://www.itu.int/ubiquitous>
13. Li, R., Li, R.-F.: A Survey of Context-Aware Computing and its System Infrastructure. Journal of Computer Research and Development 44(2), 269–276 (2007) (in Chinese)
14. Braden, R., Clark, D., Shenker, S., et al.: Developing a Next-Generation Internet Architecture. In: Proceedings of ACM SIGCOMM (2000)
15. Zhao, Y.-J., Wang, F.: Overview of Naming and Addressing Technology on Internet. Application Research of Computers 24(7), 1–7 (2007) (in Chinese)
16. Jiang, Q., He, Z.-J., Tang-lun: Technique and Prospects of Ubiquitous Network. Communications Technology (12), 181–185 (2008) (in Chinese)
17. Akyildiz, I.F.: Wireless sensor networks: a survey. Computer Networks (38), 393–422 (2002)
18. Chen, R.: Broadband Wireless Mobile Communications Development and UWB Frequency Planning in China. In: 2008 3rd CJK-WPAN/WBAM Workshop, Yokosuka, Japan (December 2008)
19. Bensley, E., Fisher, L., Gates, M., et al.: Evolvable realtime C3 systems. In: Proceedings of the 1st IEEE ICECCS, pp. 153–166. IEEE Press, Florida (1995)
20. PEO C4I and Space. Joint Airborne Network-Tactical Edge (JAN-TE) Capability Functional Description Document. China National Defence Science and Technology Information Center, Beijing (2005)
21. C4ISR AWG. C4ISR architecture framework version 2.0 (EB/OL) (December 18, 1997),  
<http://www.fas.org/irp/program/core/fw.pdf>
22. GIG Architecture Version 1.0. US:Department of Defence (July 2001)
23. DoD CIO. DoD net-centric data strategy (EB/OL),  
[http://www.afei.org/pdf/ncow/DoD\\_data\\_strategy.pdf](http://www.afei.org/pdf/ncow/DoD_data_strategy.pdf) (May 09, 2003)
24. DoD of USA. Global Information Grid Net-Centric Implementation Document: Network Operations Management (T600). Beijing: China National Defence Science and Technology Information Center (2005)
25. Ohashi, M.: Ubiquitous Network-Next Generation Context Aware Network. 2008 EU-Japan Cooperation Forum on ICT Research, Tokyo (March 2008)

# **Design and Implementation of the Patrol Terminal System for Power Transformation Facilities Based on Android and Wireless Network**

Jiantao Zhao<sup>1</sup>, Lin Bian<sup>1</sup>, and Yue Lian<sup>2</sup>

<sup>1</sup> Department of Control and Computer Engineering,

North China Electric Power University, Beijing, China

<sup>2</sup> Beijing GaoQuan Technology Co., LTD., Beijing, China

zhaojiantao66@126.com, Blin0227@gmail.com, lys39290225@163.com

**Abstract.** This paper introduces the patrol system for power transformation facilities which is needed by power industrial patrolling. It explains the design and implementation of the patrol terminal system based on Android mobile terminal development platform and wireless network, puts forward the structure model of the patrol system and expounds the design and implementation of the mobile terminal system in this patrol system. The patrol terminal system is developed on Android platform. It implements data transmission through wireless network and uses Web services to process and analyze the patrol results.

**Keywords:** Android, wireless network, Web services, patrol terminal system, power transformation facilities.

## **1 Introduction**

Along with the progress of industrial standardization, informatization and intellectualization, various kinds of intelligence information systems are constantly emerging to adapt to the needs of different industries. As the security link of safe production and normal operation, inspection has become a business which cannot be neglected by lots of industries. However, because of the different requirements and inspections of enterprise and individual, the patrolling work has a lot of uncertainty and lack of normalization. The appearance of intelligent patrol system gives the solutions to these problems to a great extent. This paper designs a kind of intelligent patrol system with data collection subsystem and data transmission subsystem developed upon Android mobile terminal development platform, in order to get wider application scope in power industry.

## **2 Key Technologies**

### **2.1 Android**

Android platform was an open software development platform developed by Google for development of embedded software used for mobile devices. It consists of five structure

compositions: Application, Application Framework, Libraries, Android Runtime and Linux kernel. The largest characteristic of Android is that it, as an open-architecture system, has the excellent environment for development and debugging, as well as the support of various scalable user experiences. Android is very rich in graphics system, has the function of multimedia support and owns strong and powerful browsers [1]. Android platform support wireless communication modes such as short message communication, Bluetooth communication, GPRS communication and Wi Fi network communication. It has a wealth of wireless communication interfaces and is very fit for software developments upon wireless network.

## 2.2 GPRS

GPRS, General Packet Radio Service, is a mobile data service which is available by GSM mobile phone users, providing mobile users with high-speed wireless IP, which can be used for Internet connection and data transmission of applications. GPRS uses the packet switching technology. It allows each user to take up several wireless channels, while every channel taken up by several users in the same time. That ensures the effective utilization of resources. Using GPRS technologies to realize data packet sent and received makes users never offline and charged according the amount of data flow, which effectively reduced the cost of services [2].

## 2.3 WLAN

The wireless local area network support communications between computers using an effective method of wireless multi-access channels. That makes the mobile communications' personalization and multimedia application possible. Popularly speaking, the Wireless local area network (WLAN), is not adopted in traditional cable at the same time, provides the functions of Ethernet or token network [3]. The wireless local area network is relevant to the IEEE 802.11 standard. It was set by large of experts in local area network and computer area, and has been ongoing perfected and edited after.

## 2.4 Web Services

Web service is a kind of service oriented to architecture technology. It provides the service through the agreed Web standard, in order to make sure the application services of different platforms can interoperability. According to the definition of the W3C, Web service should be a software system to support the interaction operations between different machines on different networks. Network services are usually consisted of many application program interfaces (API), which work through the remote servers on network such as Internet to execute service tasks required by clients [4]. The Web services in practical applications are online application services usually released by enterprises for the aim of their specific business needs. Other companies or application software can visit and use the online services through the Internet. Actually, Web service is actually a set of tools, which has a series of different calling methods. The three most common means is: remote procedure calls (RPC), service oriented architecture (SOA)

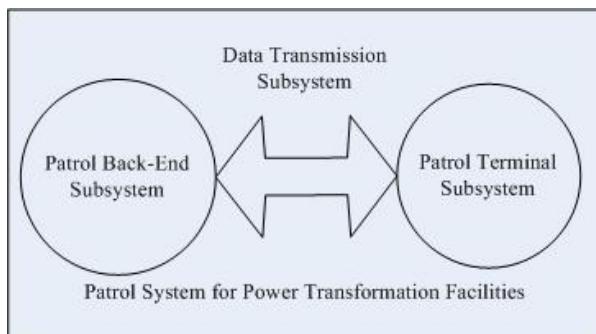
and Representational State Transfer (REST). In this paper, the Web service used was developed with Apache Axis2 and was called by Android patrol terminals with KSOAP2 tools.

### 3 System Architecture and Business Process

#### 3.1 Standardization Model

Standard is "a common used and reused normative document which is approved by recognized institutions by consensus in order to get the best order in a certain range". And standardization means the activities that people establish the rules which are common used and reused in order to solve practical problems or potential problems within a certain range.

We have come to the standardized patrol system model for power transformation facilities based on the construction, structure, process principles and other requirements defined in "Electric Power Industry Standard" for patrol system of power transformation facilities. The patrol system model is composed by patrol back-end subsystem, patrol terminal subsystem and data transmission subsystem [5].



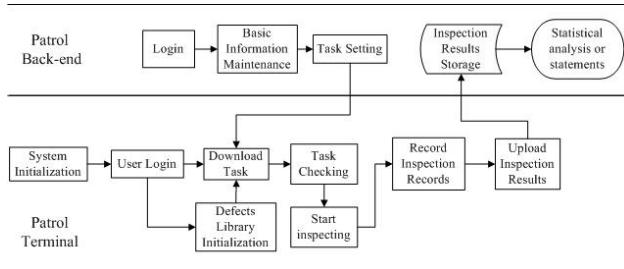
**Fig. 1.** Patrol System Architecture.

#### 3.2 Patrol System Standardization Model

The transformation facilities patrol system is divided into three parts: patrol management system references to the server-side patrol back-end system, handset side data acquisition system and data transmission system. The server-side patrol management system running on a Web server receive and process the data obtained from patrol and this section provides services to meet the needs of a variety of patrol functions, as well as secondary development interfaces; The data acquisition system uses the currently hot-used platform Android, and realizes features including download and upload tasks, personnel or mechanical identification and collecting inspection data etc., performs user-friendly and easy to record data; The data transmission system uses a combination of wireless and wired mechanism, according to the environment, the patrol

inspector can choose the way to upload or download patrol data, through a wireless or wired way.

The following diagram is the function flow of the transformation facilities patrol system.



**Fig. 2.** Function Flow Diagram

As shown in the function flow diagram, the patrol systems business process is divided into two parts, that is, the server-side operations and terminal operations. This paper focuses on the patrol terminal data acquisition system and data transmission system. These two parts sounds as the innovations of the patrol system and makes personalized improvements based on the traditional industrial PDA.

The business processes of patrol terminal can be described as follows:

- When using the terminal in the first time, the terminal should be initialized its settings, including personnel information and defect library information. When the back-end information changed, the terminal also need initialize and update.
- Inspectors login the patrol terminal system and download tasks data which should be finished.
- When the inspectors got their tasks data, they should start the patrol work and record patrol results. They can start the patrol work by sweeping the RFID of devices, or, if the RFID of device doesn't work well, they can choose the way of choosing the device manually to start the patrol work.
- When all the tasks finished, inspectors will upload the patrol results to the server. This system introduces the wireless way of transmission based on GPRS and WLAN network, as an addition to the USB cable transmission way. It works by calling the Web services for downloading and uploading data, sending the results to the patrol back-end subsystem to make analysis, on which the corresponding processing will be taken

## 4 The Design and Implementation of the Patrol Terminal System

### 4.1 Design of the Patrol Terminal System

The Patrol System for Power Transformation Facilities standard in “the Electric Power Industry Standard of People's Republic of China” (EPISPRC) makes detailed

requirements about the functions of the Patrol Terminal. The system designed in this paper tries its best to meet the requirements in the standard under the conditions that hardware allows. The basic functions of the Patrol Terminal include five main modules: system initialization, downloading patrol tasks, checking patrol tasks, starting the patrol, and uploading the patrol data. Functions of the modules are divided according to the requirements declared in the Patrol System for Power Transformation Facilities standard of EPISPRC.

**System Initialization Module:** It is used for the initialization of the patrol terminal database at the first use or when the server patrol database changes. This function module includes the functions of server IP settings and initialization of the defects database.

**Patrol Tasks Download Module:** Its function lies in obtaining patrol task data from the server. The tasks data are distinguished by ID of the staff logging in. The Authentication is made when the patrol staff logs in the patrol terminal, and the identity information is reserved in the terminal service. When downloading the patrol tasks, the system automatically selects the task corresponding to the patrol staff from the server database according to the identity information, and read as well as save the required field into the task table of the database.

**Patrol Tasks Checking Module:** It is used for checking the detailed information of the task and listing the task table, which includes the information of the patrol staff, the patrol line, and the location. Details of the task can be obtained through a detailed check.

**Starting Patrol Tasks Module:** Both automatic patrol and manual patrol are achieved in this module. The automatic patrol means after the sensing equipment of the patrol terminal get to the equipment information button, the system obtains the ID and the type of the equipment automatically as well as accesses to the patrol items automatically, which will be checked by the patrol staff item by item before input into the patrol terminal as records. As for manual patrol, the staff selects the task in the task list and selects devices in the task to patrol. The difference between the two types of patrol lies in the different trigger ways. This layer is mainly referring the enterprise information resources, to provide data services for the system.

**Uploading Patrol Data Module:** It is responsible for uploading the patrol data. The data will be stored into the terminal after the task patrol, and synchronized to the server database by calling the web services in the system.

## 4.2 The Implementation of the Patrol Terminal System

An Android application program consists of four structure blocks: Activity, Intent Receiver, Service and Content Provider, but not necessary. Beside this, it will be necessary to register the list of these structure blocks in the configuration file named `AndroidManifest.xml`. The file `AndroidManifest.xml` is used for defining the components of the application, as well as the functions and necessary conditions of these components.

The classes design process of the terminal system mainly involves three kinds of classes, that are classes for Activity, videlicet page, for Web services and for database operating.

Activity classes correspond to different pages with different functions. In Android programming, one page is called one activity, which acts as a single class inheriting to the Activity class. It realizes the communication and switchover between activities by

intents, classes inheriting to the Intent class. Take the main page class ActivityMain as an example, it implements the index page showing when users finished the authentication and logged in successfully. It contains five function buttons, which correspond to the five modules: System Initialization Module, Patrol Tasks Download Module, Patrol Tasks Checking Module, Starting the Patrol Module and Uploading the Patrol Data Module. The class ActivityMain defines the operations of these buttons. If triggering the event will cause a skip to other activity, it will be necessary to get an intent instance to call the activity to skip to. Relevantly, if there is no skip to other activities, it will not involve an intent instance.

There are three classes for Web services: SoapObject class, SoapSerializationEnvelope class and HttpTransportSE class. The system draws support from the open source framework KSOAP2 to develop web services for patrol terminal. These three classes reference the class definitions in KSOAP2. The SoapObject class defines that the request and its parameters of Web service be encapsulated in SoapObject instance. SoapSerializationEnvelope class performs as an envelope with SOAP protocol in it, encapsulating SoapObject objects inside. HttpTransportSE class describes the Web service communications based on SOAP protocol meaning to transmit data through HTTP protocol, encapsulating SoapSerializationEnvelope objects inside.

Database operation class points to the DataBaseHelper class, which inherits the SQLiteOpenHelper class in Android SDK, defining the operations of database. Applications developed on Android usually use the open source database SQLite. It takes up little memory space and enough to meet the desire of application usage. SQLiteOpenHelper class goes in accordance with the Object-Oriented thoughts, defines the operations that applications do to the databases. DataBaseHelper class implements and advances on the base of SQLiteOpenHelper class, to make the database operations meet the actual needs in the system.

### **4.3 The Implementation of Communication Function**

This patrol system chooses the wireless way to complete data communication, and Android, as a mobile terminal development platform, has innate superiority of supporting wireless network. So long as the chosen patrol device has the corresponding hardware support, realizing the wireless way of data communication with GPRS, 3G or WLAN will be not too difficult. In order to reduce the cost of development, the patrol system did not choose to develop hardware module from the bottom level, but completed the communication function of terminal with top level protocol HTTP in wireless network environment. Android provides rich interfaces for wireless network, and the patrol system uses its GPRS and WLAN interfaces to communicating, and raises different ways of data processing aiming at different network environments [6].

Given the problem of authority limits between different databases, the patrol terminal cannot modify data in server database SQL Server directly; as a result, the system uses Web service to carry out data communication between patrol terminals and patrol server. The server announces a Web service project with Apache Axis2 agent, and the terminal system calls the Web service with the KSOAP2 method to communicate. In the Web service implement, the data to transmit will be packaging in a SoapObject object firstly,

after transmitted in the way of data stream, it will be formed as a SoapObject object again, then be operated with SQL statements stored in the database.

Communication between the terminal and the server is mainly used for four places: user authentication, initialize the defect library, download patrol tasks and upload patrol results. Web service provides methods for remote authentication and there are no data transfers. Initialized defect library is transferred by whole table. The database table of defects library is transferred to the patrol terminal and stored in the terminal database by the web service method which is defined in the server. SQL statements are used to select the corresponding task records and then the data can be downloaded and stored in the terminal database. When the terminal uploads the patrol data, not all the fields of records need to be modified. Data to send back to server will be compared with the original data in server using SQL statements, if the data were the same as original, there will be no need to transmit it back. Only the data which has changed will be sent back to server. This approach reduces the amount of data transmission and the server database modification and improves security and stability.

## 5 Verification of Correctness

After coding and testing, the system can be operated in the proposed way. It can meet the needs of the normal inspection mission and can take the data transmission through wireless network.

In the transmission function aspects, the system, with the Web service, can complete the table download. Defect library table initialization ensures the required inspection items be automatically taken for inspection. When uploading inspection results, the system packages the data to upload in the form of a string, and then transform it into a stream for transmission. In the server side, the stream is turned into a string which will be parsed and stored in the database.



**Fig. 3.** System main interface functions

## 6 Conclusion

This paper introduces the patrol system for power transformation facilities which has been needed by power industrial patrolling. It explains the design and implementation of the patrol system based on Android mobile terminal development platform and wireless network, puts forward the structure model of the patrol system and expounds the design and implementation of the mobile terminal system in this patrol system.

## References

1. Di, S.: Android revelation and future development tendency (EB/OL),  
<http://publish.itpub.net/zt/android/index.html> (March 17, 2008)
2. Wen, Z.: General packet radio service – GPRS. Electronic Industry Press, Peking (2004)
3. The wireless local area network technology,  
[http://www.grchina.com/tech\\_wireless.htm](http://www.grchina.com/tech_wireless.htm)
4. Yang, F.: Android application development revelation, pp. 41–47. China Machine Press, Peking (2010)
5. Ball, J., Carson, D.B., Evans, I., Haase, K., Jendrock, E.: The Java EE 5 Tutorial. Sun Microsystems, Santa Clara, pp. 42–46 (2006)
6. Ma, Y.: The structure of the Android applications. Peking: China university of geosciences (Beijing) master's degree thesis, pp. 9–20 (2008)

# Dynamic One-Way Key Establishment Scheme in Wireless Sensor Networks

Sisi Jiang<sup>1</sup>, Zhengye Si<sup>2</sup>, Zushun Wu<sup>1</sup>, and Dan Li<sup>1</sup>

<sup>1</sup> Department of Electronics, School of Information Science,  
Beijing Normal University, Beijing, China

<sup>2</sup> Department of Mathematics, School of Science,  
Harbin Institute of Technology, Weihai, China

{jiangsisino1, zkjy2006, wuzs100, lidan}@126.com

**Abstract.** The existing key establishment scheme in wireless sensor networks cannot prevent the adversary from discovering more and more keys of the networks, and thereby, after a long time, the network would be controlled by the adversary. Besides, these schemes also cannot provide a sound method to prevent the replica attacks. In this paper, we propose a dynamic one-way key establishment scheme (called DOK for short hereinafter) which introduces a dynamic one-way key strategy to prevent the adversary from discovering more keys and posing the replica attacks to the networks.

**Keywords:** Wireless sensor network, key establishment scheme, replica attacks, DOK.

## 1 Introduction

Eschenauer and Gligor proposed the basic random keys scheme in [4], where a large key pool K is computed offline and each sensor picks k keys randomly from K without replacement to form a key ring before deployment. Only if two sensors have at least one key in common they could communicate with each other securely. In literature [6], an enhanced scheme of the basic random keys scheme is proposed in which if two nodes want to communicate with each other, they should share more than  $q > 1$  number of common keys. However, the randomness characteristics caused by the unavailable deployment information in wireless networks forces these schemes using up more storage space for the key information to guarantee the key-sharing probability. Recently, some excellent key establishment schemes are proposed such as LKE (A Self-Configuring Scheme for Location-Aware Key Establishment in Wireless Networks) [5], and it belongs to the location-aware key establishment schemes. LKE not only could scale well to large-scale sensor networks but also is able to provide good security performance, high key-sharing probability and low storage overhead. However, the keys produced in literatures [4], [5], and [6] are all static. Besides, they all did not introduce an effective solution for the replica attacks. In order to distinguish these schemes from our DOK, we call them as static key establishment

scheme. DOK takes advantage of the dynamic one-way key strategy to stop the adversary getting more keys of the sensor network. In addition, because the DOK eliminates the randomness characteristics, it could provide a higher key-distributed probability and lower storage overhead.

## 2 Network Models

In order to avoid the local broadcasting collision, a reliable MAC protocol proposed by literature [1] is introduced. We assume that the clocks of all nodes could be synchronized by employing some secure time synchronization protocols e.g. [2], [3] such that all the sensor nodes can trigger a new round of key establishment procedure at the same time.

## 3 The DOK Scheme

The discussion about our DOK scheme includes two parts: the key establishment and the strategy against coming-back nodes and replica attacks. The detailed content of the DOK is as follows.

### 3.1 Key Establishment

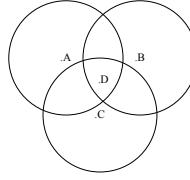
The first part includes three phases, and they are as follows.

#### 1) Pre-establishment

A key  $k_0$ , a random number generation function  $R$ , a key generation function  $F$  and a key updating algorithm are preloaded to all the network sensors. The initial secret key  $k_0$  is preloaded in all the sensor nodes such that all the messages exchanged can be protected during the first round key establishment procedure. Therefore  $k_0$  should be strong enough so that it is almost impossible for an adversary to recover it before finishing the first round of key establishment procedure. For each round of key establishment, a random number is generated by the function  $R$  (the CPU occupancy rate, etc.), and input to the function  $F$  to generate a key for the current communication round. Then the key updating algorithm is executed to perform a key updating procedure which will be discussed in detail below. Besides, we should divide Time Domain into lots of time epochs of equal length “ $\Delta t$ ”. At the beginning of each  $\Delta t$ , a new round of key establishment process will be triggered, and then the keys of all the nodes will be updated.

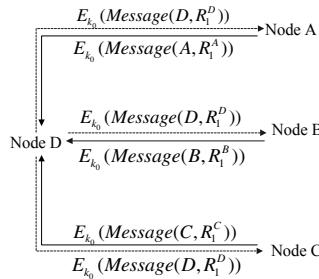
#### 2) The first round of one-way key establishment procedure

After the deployment of a sensor network, each node is expected to exchange its random number with its neighbors and these neighbors will use this random number to compute their secret keys. In order to explain this process in detail, we take nodes  $A$ ,  $B$ ,  $C$  and  $D$  in Figure 1 for example.



**Fig. 1.** Node  $D$  and its neighbors  $A$ ,  $B$  and  $C$

As soon as the key establishment procedure is triggered, all the sensor nodes are expected to exchange the message  $Message(ID, R_i^{ID})$  with their neighbors, where  $R_i^{ID}$  is the random number for the  $i$  round of key establishment. For example, node  $D$  sends the message  $Message(D, R_1^D)$  to  $A$ ,  $B$ ,  $C$  for the first round of key establishment, where  $D$  is the ID of node  $D$  and  $R_1^D$  is a random number generated by the function  $R$  for the first round of key establishment. Simultaneously, nodes  $A$ ,  $B$ , and  $C$  send  $Message(A, R_1^A)$ ,  $Message(B, R_1^B)$ ,  $Message(C, R_1^C)$  to node  $D$  respectively. All of these messages are encrypted by  $k_0$ . The message exchange process is shown in Figure 2.



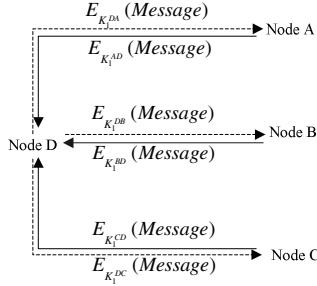
**Fig. 2.** The message exchange process between  $D$  and its neighbors  $A$ ,  $B$  and  $C$

After the message exchange process, node  $D$  can compute its keys for the first round with  $F$  as follows:

$$\begin{aligned} K_1^{AD} &= F(R_1^A, R_1^D), \quad K_1^{DA} = F(R_1^D, R_1^A) \\ K_1^{BD} &= F(R_1^B, R_1^D), \quad K_1^{DB} = F(R_1^D, R_1^B) \\ K_1^{CD} &= F(R_1^C, R_1^D), \quad K_1^{DC} = F(R_1^D, R_1^C) \end{aligned}$$

We should choose the key generation function  $F$  which is able to make sure that  $F(x, y) \neq F(y, x)$ . That is to say,  $K_1^{AD} \neq K_1^{DA}$ ,  $K_1^{BD} \neq K_1^{DB}$ ,  $K_1^{CD} \neq K_1^{DC}$ . Similarly,  $A$  computes  $K_1^{AD}$  and  $K_1^{DA}$ ,  $B$  computes  $K_1^{BD}$  and  $K_1^{DB}$ , and  $C$  computes  $K_1^{CD}$  and  $K_1^{DC}$ . As a result,  $A$  shares  $K_1^{AD}$  and  $K_1^{DA}$  with  $D$ ,  $B$  shares  $K_1^{BD}$  and  $K_1^{DB}$  with  $D$ , and  $C$  shares  $K_1^{CD}$  and  $K_1^{DC}$  with  $D$ . Besides, as mentioned above, all the keys in our DOK scheme could only work on one way. For

example,  $K_1^{DA}$  could only be used by node  $D$  to encrypt all the messages sent to  $A$  while  $K_1^{AD}$  could only be used by node  $D$  to decrypt the messages sent from  $A$ , so  $K_1^{DA}$  and  $K_1^{AD}$  are one-way keys for  $D$ . This process is shown in Figure 3, where the  $E_{K_1^{DA}}(Message)$  is a cipher text which is the message encrypted by key  $K_1^{DA}$ .



**Fig. 3.** The communication process between  $D$  and  $A, B, C$

### 3) The key updating process

As mentioned in the phase 1, all the nodes in the network should update its keys at the beginning of each  $\Delta t$ , and the key updating procedure could be carried out by the key updating algorithm which is described as follows:

---

#### Key Updating Algorithm (executed in the source node )

---

**Input:**  $F, R, K_i^{Sor-Des}, K_i^{Des-Sor}, R_{i+1}^{Des}, SorID, DesID$

**Output:**  $K_{i+1}^{Sor-Des}, K_{i+1}^{Des-Sor}$

**1:**  $R_{i+1}^{Sor} = R(i+1)$ ; //Source node generates a new random number  $R_{i+1}^{Sor}$ .

**2:** Send message  $E_{K_i^{Sor-Des}}(SorID, R_{i+1}^{Sor})$  to its neighbors;

**3:** Receive message  $E_{K_i^{Des-Sor}}(DesID, R_{i+1}^{Des})$  from its neighbors;

**4:** Decrypt message  $E_{K_i^{Des-Sor}}(DesID, R_{i+1}^{Des})$  with  $K_i^{Des-Sor}$  and get  $DesID$  and  $R_{i+1}^{Des}$ ;

**5:**  $K_{i+1}^{Sor-Des} = F(R_{i+1}^{Sor}, R_{i+1}^{Des})$ ;

**6:**  $K_{i+1}^{Des-Sor} = F(R_{i+1}^{Des}, R_{i+1}^{Sor})$ ;

**End function**

---

For understanding this algorithm better, we take nodes  $D$  and  $A$  for example, and  $D$  is the source node and  $A$  is the destination. At the beginning of  $\Delta t$ ,  $D$  would generate a random number  $R_{i+1}^D$  with function  $R$ , and then send its own ID and  $R_{i+1}^D$  encrypted by  $K_i^{DA}$  (i.e.  $E_{K_i^{Source-Des}}(SourceID, R_{i+1}^{Source})$ ) to  $A$ . At the same time,  $D$  receives the message  $E_{K_i^{AD}}(A, R_{i+1}^A)$  sent by  $A$ . Since  $D$  has  $K_i^{AD}$ ,  $D$  can decrypt  $E_{K_i^{AD}}(A, R_{i+1}^A)$

and then get  $A$  and  $R_{i+1}^A$ . After that,  $R_{i+1}^A$  could be used to compute  $K_{i+1}^{AD}$ . Similarly,  $A$  could compute  $K_{i+1}^{DA}$ . For the next round communication,  $D$  and  $A$  would use the new keys to communicate with each other safely.

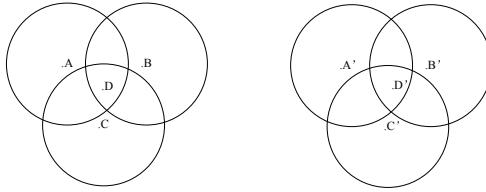
### 3.2 A Strategy against Coming-Back Nodes and Replica Attacks

In this part, we propose a strategy of dealing with the coming-back nodes and the replica attacks. Since keys in all the nodes are updated again and again, if a node quits the sensor network for a period which is longer than the interval  $\Delta t$ , this node cannot join into the communication again according to the strategy mentioned above. That is to say, when this node resumes, it has no current key information, so it cannot come back the network and work as usual. In order to cope with this problem, we present a strategy for the coming-back nodes. Besides, the strategy is also effective for defeating the replica attacks. This strategy is described as follows:

- Each node in the network informs all the neighboring information to its neighbors, so each node knows the nodes within two-hop distance away.
- If a node quits for a period owing to some reason, when a new key updating round begins, all the neighbors of this node would use the random number sent by this node before quitting the network and the new random numbers generated by them to compute new keys. If this node comes back, firstly he should notify his neighbors that he wants to join the network, then all his neighbors would check whether he is their collective neighbor.
- If yes, these neighbors would give the current random numbers to this node and this node could compute the new keys with these random numbers sent by his neighbors and the random number stored before he quits, then the node could come back the network. Else, he cannot be accepted to join the network.

In order to understand this strategy better, we also take nodes  $A$ ,  $B$ ,  $C$  and  $D$  in Figure 1 for example. The locations of  $A$ ,  $B$ ,  $C$  and  $D$  are as shown in Figure 1. Node  $D$  would inform its neighboring information to  $A$ ,  $B$  and  $C$ . That is to say,  $A$ ,  $B$  and  $C$  know the existence of each other. We assume that before  $D$  quits the network,  $A$ ,  $B$ ,  $C$  and  $D$  use random number  $R_i^D$  to generate  $K_i^{AD}, K_i^{DA}, K_i^{BD}, K_i^{DB}, K_i^{CD}, K_i^{DC}$ . Then after  $D$  quits, nodes  $A$ ,  $B$  and  $C$  would use  $R_i^D$  and  $R_j^A, R_j^B, R_j^C (j > i)$  to generate new keys  $K_j^{AD}, K_j^{DA}, K_j^{BD}, K_j^{DB}, K_j^{CD}, K_j^{DC}$ . When  $D$  comes back, nodes  $A$ ,  $B$  and  $C$  would communicate with each other to check whether  $D$  is their collective neighbor. If yes, nodes  $A$ ,  $B$  and  $C$  would give  $D$  the current  $R_k^A, R_k^B, R_k^C (i < j \leq k)$ , and then  $D$  could compute the current keys  $K_k^{AD}, K_k^{DA}, K_k^{BD}, K_k^{DB}, K_k^{CD}, K_k^{DC}$  using  $R_i^D$  and  $R_k^A, R_k^B, R_k^C$ . If not, nodes  $A$ ,  $B$  and  $C$  would refuse to communicate with  $D$ . In this way, a node who quits the sensor network could come back again.

In addition, our DOK can also defeat the replica attacks using this strategy. We assume that  $D'$  is the replica of  $D$ , replica node  $D'$  is deployed randomly in the network, and  $A', B', C'$  are the neighbors of  $D'$ , as shown in Figure 4.



**Fig. 4.** The deployment of a replica node  $D'$  in the network

If node  $D'$  wants to communicate with  $A', B', C'$ , nodes  $A', B', C'$  would check whether  $D'$  is their collective neighbor. Because  $D'$  is an unknown node to  $A', B', C'$ , nodes  $A', B', C'$  would refuse to communicate with  $D'$ . As a result, node  $D'$  cannot establish keys with other nodes, either. Therefore, our DOK can defeat the replica attacks.

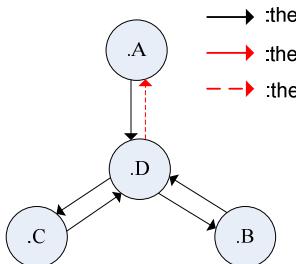
## 4 Evaluation of the DOK Scheme

In this section, the DOK scheme is evaluated in terms of security and the storage, computation and communication overheads.

### 4.1 Security

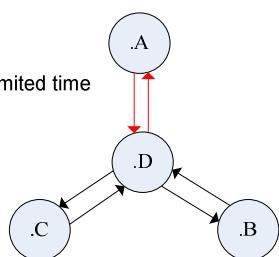
Since the location of each network node is immobile after deployment, the probability that two neighboring sensors are able to establish a key can almost reach to 100%. So it is obvious that our DOK can guarantee the key distributed probability.

We firstly discuss the situation when the adversary can only guess the keys depending on the captured messages. Because each node uses one key to encrypt messages and uses the other key to decrypt messages, we call these keys as one-way keys. For instance, node  $D$  uses  $K_{DA}$  to encrypt all the messages he sends to  $A$  and uses  $K_{AD}$  to decrypt the messages sent to him by  $A$ . That is to say,  $K_{DA}$  and  $K_{AD}$  could only work on one way by the node  $D$  or  $A$ . If an adversary discovers  $K_{DA}$ , then he could only get the messages sent from  $D$  to  $A$ . However, owing to the dynamic feature of the DOK scheme, this adversary could only use  $K_{DA}$  for a limited period of time. So comparing with the static key scheme, our DOK not only could limit the number of attacked links but also could limit the time that an adversary uses the discovered keys, and these situations are shown in Figure 5.



**Fig. 5(a).** Affected links in DOK

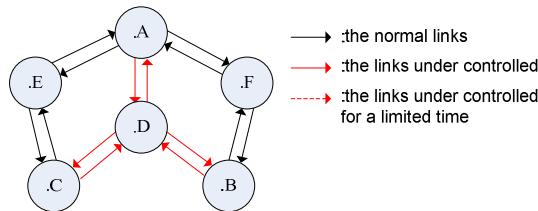
→ :the normal links  
 → :the links under controlled  
 → :the links under controlled for a limited time



**Fig. 5(b).** Affected links in others

From Figure 5, we can see that if one key is discovered by an adversary, only a one-way link between two nodes will be invaded in the DOK while in other schemes the whole link would be controlled by the adversary. In addition, the adversary could only use this discovered key for a limited time for DOK while in other schemes the discovered keys could be used all the time by the adversary.

Secondly, we pay attention to a more serious situation when the adversary captures a node, gets all the key information stored in this node, and then poses a replica attacks on the sensor network. Although the adversary captures a node, obtains all the key information stored in the node, and then poses a replica attacks on the network, our DOK still can provide an excellent security performance. We take nodes A, B, C and D in Figure 5 for example. If node D is captured, then the adversary could get  $K_i^{AD}, K_i^{DA}, K_i^{BD}, K_i^{DB}, K_i^{CD}, K_i^{DC}, F$  and  $R$ . Comparing with the other key establishment schemes such as LKE, the affected links in the DOK scheme is the same as other schemes (see Figure 6).



**Fig. 6.** The status of links under controlled in DOK and LKE schemes

From Figure 6, it can be seen that the network links between  $D$  and  $A, B, C$  are all under controlled, and because the adversary knows the functions  $F$  and  $R$ , the dynamic strategy in our DOK would be not effective any more. However, the security performance of the sensor network served by the DOK scheme is the same as that of the other schemes. Even more, the DOK introduces a strategy to defeat the replica attacks so that the security of the network is further enhanced. Consequently, we can say that our DOK could provide a better security performance than the existing key establishment schemes.

## 4.2 Storage, Computation and Communication Overheads

Comparing with the key establishment scheme using pre-loaded key space, such as the basic random-keys scheme [6], our DOK eliminates the randomness characteristics, and therefore doesn't need to store lots of keys to guarantee the key-sharing probability, so the storage space required by each sensor must be smaller. Moreover, comparing with some self-configure key establishment scheme such as LKE [5], etc., we also save the storage space to some extent. For LKE, the storage overhead is  $\pi \times (\lambda + 1) \times \log s'$  and  $\lambda = \pi L^2 N / A$  where  $L$  is the grid size which determines the coverage area of a server sensor and this area is expected to cover  $\lambda$  worker sensors, so the storage overhead is  $m_2 = \pi \times (\pi L^2 N / A + 1) \times \log s'$ .

We could have  $m_2/m_1 \approx \pi L^2 \log s' / 2r^2 s$ . For example, suppose the length of a key is 128, then  $s=128$  and  $\log s' = \log(2^{128}) \approx 39$ , so there is  $m_2/m_1 \approx (\pi L^2 \times 39) / (r^2 \times 256) \approx 0.2 \times (L/r)^2$ . It is well known that  $L \gg 2.23r$ , so  $m_2/m_1 \gg 1$ , that is to say, the storage overhead of our scheme is remarkably small.

Because our DOK use the dynamic key strategy to improve the security performance, the communication and computation overhead would increase along with time. Therefore, our communication and computation overhead must be larger than the static key establishment scheme. Our future work would concentrate on how to balance the security demand and the computation overhead.

## 5 Conclusion

In this paper, we propose a DOK scheme, a dynamic and one-way key establishment scheme for wireless sensor networks. The performance evaluation shows that some outstanding capabilities could be achieved, such as excellent security performance, high key-sharing probability, low storage expense, and in the mean time the computation and communication overhead is a bit larger rationally. For future work, we will try to design an algorithm to balance the security and the overhead. Icular paper. Digital signatures are acceptable.

## References

1. Ye, W., Heidemann, J., Estrin, D.: An Energy-Efficient MAC Protocol for Wireless Sensor Networks. In: Proc. IEEE INFOCOM 2002, pp. 1567–1577 (2002)
2. Song, H., Zhu, S., Cao, G.: Attack-resilient Time Synchronization for Wireless Networks. Ad Hoc Networks. 112–125 (2007)
3. Sun, K., Ning, P., Wang, C., Liu, A., Zhou, Y.: TinySeRSync: Secure and Resilient Time Synchronization in Wireless Networks. In: ACM CCS, Alexandria, pp. 264–271 (2006)
4. Liu, D., Ning, P.: Establishing Pairwise Keys in Distributed Sensor Networks. In: The 10th ACM Conference on Computer and Communications Security (CCS 2003), pp. 52–61 (2003)
5. Liu, F., Cheng, X.: LKE: A Self-configuring Scheme for Location-aware Key Establishment in Wireless Sensor Networks. IEEE Transactions on Wireless Communications 224–232 (2008)
6. Chan, H., Perrig, A., Song, D.: Random key pre distribution schemes for sensor networks. In: Proc. S&P 2003: the 24th IEEE Symposium on Security and Privacy, pp. 197–215 (2003)

# Query-Aware Location Privacy Model Based on p-Sensitive and k-Anonymity for Road Networks

Jiahui Chen, Hongyun Xu, and Lin Zhu

School of Computer Science and Engineering, South China University of Technology,  
GuangZhou 510006

**Abstract.** Recently, several techniques have been proposed to protect users location privacy for location-based services in the Euclidean space. Applying these techniques directly to the road network environment would lead to privacy leakage and inefficient query processing. In this paper, we propose a query-aware location privacy model based on p-sensitive & k-anonymity for road networks. By cloaking a user's location into  $p$  connected road segments which include at least  $k$  users, our model is able to meet privacy requirements, and to minimize the query processing cost. We conducted simulate experiments on our model, the result manifest its privacy resilience and efficient query processing.

**Keywords:** Location privacy, location-based services, p-sensitive & k-anonymity.

## 1 Introduction

Location-based services (LBS) combine the functionality of location-aware devices (e.g., GPS devices), wireless and cellular phone technologies, and information management, to provide personalized services to users based on their locations. Examples of LBS include location-aware emergency services, e.g. "Dispatch the nearest ambulance," location-based advertisement, e.g. "Send e-coupons to all cars that are within two miles of my gas station," live traffic reports, e.g. "Let me know if there is congestion within ten minutes of my route," and location-based store finders, e.g. "Where is my nearest restaurant?". In this context, users registered to LBS and continuously send their locations to a location-based database server. Thus, the users suffer from privacy leakage by disclosing their location to un-trustworthy servers. In order to protect location privacy in LBS, several techniques have been proposed to anonymize the user's location [1,2-5,6,9]. Nevertheless, assuming the random-waypoint mobility model, wherein users can move in arbitrary directions at random speed, these existing solutions fail to address the vulnerabilities of mobile users traveling over roads, where both the user mobility and the location-based service processing are constrained by the underlying road network environment. What's more, although some techniques have been proposed to preserve the user location privacy for location-based services in road network environments [6,8], they have different limitations (see section 2).

In this paper, we focus on the spatial cloaking techniques in the road network which not only are able to satisfy user's privacy requirements (e.g.  $k$ -anonymity and  $p$ -sensitive), but also are able to trade-off between the query processing cost and the query quality. Additionally, our model is able to minimize query processing cost.

To sum up, our contributions are as follow:

1. Our technique can efficiently protect privacy of the query user in the road network environment.
2. Our technique can balance between query processing cost and query quality, additionally can ensure the best query processing cost, which improve the expansibility of the LBS.
3. Our definition of user privacy requirements includes  $k$  and  $p$ , which guarantees personalized  $k$ -anonymity level and  $p$ -road network segments, is a more user-oriented definition.

The rest of the paper is organized as follows: Section 2 overviews related work. Section 3 describes our system architecture. Section 4 presents our analysis model and proposes our algorithms for simple greedy and random greedy, respectively. Section 5 depicts a novel and expandible shared execution schema. Section 6 evaluates the proposed techniques with comprehensive experiments, and Section 7 concludes the paper with a discussion.

## 2 Concepts and Models

In this section, we give the concepts of this paper, define the user requirements, and present our system model and privacy attack model.

### 2.1 Concepts

To propose the mobile users/privacy requirements, we summary some definition in the paper [7, 8], and give the definitions as follow.

*Location  $k$ -anonymity: A user's reported location is said to be  $k$ -anonymous, if at least ( $k-1$ ) other active users report the same location.*

*Segment  $p$ -sensititation: A user's published location is said to be  $p$ -sensititation, if it satisfies location  $k$ -anonymity, and contains at least  $p$  different road segments.*

According to these definitions, we set our user/privacy requirement as defining  $k$  and  $p$ .

### 2.2 Privacy Attack Model

Since our system only preserves the user location privacy for snapshot location-based queries, we assume that an adversary is unable to infer that some particular snapshot queries are issued by the same user or track a particular user. Beyond this assumption, we describe two attack models, namely, replay attack [6, 7] and center-of-cloaked-attack [6, 9].

### 3 Efficient Balanced Query Algorithms

In this section, we first describe query processing cost and query quality. Then, we describe how to balance between the query processing cost and query quality under the user's privacy requirements. Finally, we give our efficient balanced query algorithms.

#### 3.1 Query Processing Cost and Query Quality

As we all know, there are two main factors that control the quality of a cloaked set of road segments  $S$ , namely, the query processing cost and the query quality, respectively. In order to minimize the query processing cost, we will need to select an  $S$  (cloaked segment set) that satisfies the user privacy requirements while optimizes the query processing cost. It has been proved that the query processing cost on the road network environment is the sum of the execution cost of the range search and external search steps [7]. The range search step indicates when giving a cloaked segment set, the LBS have to execute a range query for the whole cloaked segment set, and return a result set as candidate set. Then, the LBS have to execute another external query for every open node. Let  $C_s$  and  $C_n$  denote the computation cost (in terms of both CPU and IO) of a segment and an open node, respectively. Hence, for a typical query with a set of segments  $S$ , the processing cost,  $\text{Cost}(S)$ , can be approximately estimated as follows:  $\text{Cost}(S) = C_n \cdot V_o(S) + C_s \cdot E(S)$ , where  $V_o(S)$  and  $E(S)$  denotes the number of open nodes and the number of segments in the giving cloaked segment set, respectively. Since in our case, we denote the user requirement  $p$ -road network segments, so the  $E(S)$  may approximate to  $p$ , we can give a conclusion that the influent factor of the query processing cost is  $V_o(S)$ , which means the number of open nodes in the cloaked segment set.

On the other hand, to optimize the query quality, we need to select an  $S$  that satisfies the user privacy requirements, and has the number of users as close as possible to the anonymity  $k$  requirements and the shortest length. The main idea is that the shorter the length of  $S$ , or the less number of users of  $S$ , the smaller the size of the candidate list [7], and hence the better query quality will provide.

#### 3.2 Balance Analysis Model

Due to the analysis above, we are trying to get a balance between query processing cost and query quality. Since the main influent factor of query processing cost is the number of open nodes,  $V_o(S)$ , to obtain the optimal query processing cost, we introduce a list of node degree, namely, `NodeDegree_list`, which shows the node identifier and the node degree in cloaked segment set  $S$ . We may use the following function to select a best query processing cost for our anonymization processing.

```

Function 1 QCost(S,Q)

1: R<-the segment contains the user U
2: put all nodes in S into NodeDegree_list and set node degree
3: for every node i in NodeDegree_list do
4:     S'<-the segment not in S but contains node i
5:     if NodeDegree_list.nodei.degree>1 then
6:         NodeDegree_list.nodei.degree++
7:         insert another node into NodeDegree_list
8:     remove from R the segment contains the user U
9:     R<-S'.segment
10:    end if
11: end for
12: return R

```

The algorithm has two inputs, S and Q, It first puts all nodes in S into NodeDegree\_list, and secondly for each node i in NodeDegree\_list, it searches all the segments not in S but contains node i as candidate segments, then the algorithm checks if its degree is bigger than 1, if so, the algorithm returns the segment R contains the current target.

On the other hand, the query quality is measured by the number of candidate target objects returned by a database server given that the query location is within a cloaked set of road segments S. For the  $k$ -anonymity privacy requirement, we consider two cases. Case 1: When a current cloaked set of segments S has not yet satisfied this privacy requirement, to minimize the number of segments and satisfied this requirement, we should select some segments that contain more users to S. And we set the influence parameter in this case as  $\text{NumUser}(S)/k$ . Case 2: if S satisfied the  $k$ -anonymity privacy, since the number of user is meaningless for the anonymization processing, we set the influence factor of a segment in this case as 1. For the  $p$  privacy requirement, since our algorithm will add one segment every time, we should select the segment has shortest length. So we set the influence factor in the p requirement as  $1/\text{Length}(S)$ . Hence, the combined query quality function is:

$$\begin{aligned} \text{Quality}(S) &= \text{NumUser}(S) / k * 1 / \text{Length}(S), \text{ if } \text{NumUser} < k \\ &= 1 / \text{Length}(S), \text{ if } \text{NumUser} \geq k \end{aligned}$$

### 3.3 Simple Greedy Algorithm

The main idea of our simple greedy algorithm is to search from the road segment contains the query user, namely the user segment, and add it to the cloaked set S, if this segment satisfy the privacy requirement, it simply return S to the server. Otherwise, it

first does the best query processing cost selection, which returns a best segment (or some best segments) from all the adjacent road segments of S, if the algorithm could not find a best query processing cost segment, it will select the segment with best query quality from all adjacent road segments of S, then the algorithm add the selected segment to S. the algorithm may loop until S satisfy the privacy requirement.

Algorithm 1 depicts the pseudo code of our simple greedy algorithm. The algorithm has two input parameters, the identifier of the user U who issues the query and the issued query Q, and it will return an S.

```
Algorithm 1 Simple_Greedy (user U, Query Q)

1: e<-the road segment contains the user U
2: S<-{e}
3: while NumUser(S)<U.k or NumSegment(S)<U.p do
4:         R<-All adjacent road segments of S
5:         R'<-Qcost(S,Q)
6:         if the number of segments of R > 1 then
7:             BestSegment<-argmax ei•R'(Quality(S+ei,Q))
8:         else
9:             BestSegment=Qcost(S,Q)
10:            if BestSegment equal to e then
11:                BestSegment<-argmax ei•R(Quality(S+ei,Q))
12:            end if
13:        end if
14:        S<-S+BestSegment
15: end while
16: return S
```

### 3.4 Random Greedy Algorithm

In this section, we notice our simple greedy algorithm is vulnerable to attacker, especially the replay attack. So we inject some random parameter [6] to the simple greedy algorithm, to make it uncertain. In our random greedy algorithm, we supply two random factors, namely, randF and randN, to control our selection of segment.

In our random greedy algorithm, the algorithm first set a random factor RanF  $\in \{0,1\}$ , when the algorithm needs to select a segment from adjacent segment set R, the algorithm randomly generated another random factor RanN  $\in \{0,1\}$ , if RanN > RanF, it randomly selected an segment in R, otherwise it chose a best segment of R using the simple greed method.

## 4 Shared Execution Schema

As a location-based database server is likely to receive a numerous number of concurrent queries in the same time, processing these queries individually would pose a system bottleneck. In this section, we propose a novel shared execution schema to minimize the number of queries executed by the location-based database server for a set of private queries in the same time. Our motivation of this shared execution schema is that two or more queries can share if they belong to the same query type and interested in the same candidate result set. our shared execution algorithm is achieved by setting priorities, where we set the road segment that contains each user the highest priority, and all segments of the query users is added to the cloaked area S, then we place each user's segment into  $S_i$ , if S can satisfy all users requirements, is returned. Secondly, we consider each user's adjacent segments of  $S_i$  sets  $R_1, \dots, R_n$  the second priority. In the second priority, we first select the most frequent segments beyond  $R_1, \dots, R_n$ , and if these segments are more than 1, we use our simple greedy algorithm to select an optimal segment. Finally, we set all adjacent segments of S the third priority, and used the same simple greedy algorithm to choose a best segment to add to S. The algorithm ended when S satisfy all the query users requirements.

## 5 Experimental Results

In this section, we perform an empirical analysis of the location anonymization model proposed in this paper, which designed and programmed for the algorithms of our paper and the paper [6] used Java programming. Due to our user requirement includes  $p$ , we named our simple greedy algorithm and random greedy algorithm PPG and PRG, respectively. Similarly we named the two algorithms of paper [6] LPG and LRG, respectively.

In all experiments, we generate a set of moving objects on the road map of Oldenburg, using the Network-based Generator of Moving Objects by T. Brinkhoff, a state-of-the-art traffic simulator. The road map has 6,105 segments and 7,035 nodes in which the average length of the segments is 184m. Mobile users are initially distributed among the nodes, and then move along the roads at speeds 50 miles per hour. The experiments were run on a Windows XP system with an Intel Pentium Dual-Core processor at 2.20GHz and 2GB RAM.

**Parameter Settings.** Unless mentioned, the experiments consider 20,000 mobile users in the underlying road network in which 1,000 users issue queries at the same time. The random factor RandF for our randomized greedy approach is set to 0.2. Since we use our algorithm to compare with algorithms in the paper [16], and taking into account the average length of the segments of the road map is 184m, we set our default user requirements is  $k = 20$ ,  $p = 6$ . And the baseline paper [16] user requirement is  $k = 20$ ,  $L = 1000$ .

**Table 1.** Parameter settings

Parameter	Default value	Evaluation Range
Number of users	20000	10000-50000
Number of queries	1000	600-1000
Anonymity levels ( $k$ )	20	10-30
Minimum segment number ( $p$ )	6	5-10
Minimum segment length ( $L$ )	1000	800-1800
Random factor (ranF)	0.2	0.1-0.5

**Performance Metrics.** We evaluate our algorithms with respect to four performance measures, (1) the anonymization time, (2) the number of open nodes, (3) the success rate of the replay privacy attack, and (4) the success rate of the center-of-cloaked-area privacy attack. The anonymization time is the average time consumed in the anonymization processing, it shows the anonymous costs in the location anonymizer. The number of open nodes in the cloaked segments set  $S$  is the average statistical number of open nodes in  $S$ , it shows the query processing cost and query quality of the LBS giving a certain cloaked segment set  $S$ . The success rate of the two privacy attacks indicates the privacy resilience of our algorithms.

## 5.1 Number of Users

Figures 1 depict the scalability of our approaches with respect to varying the number of users from 10,000 to 50,000.

Figure 1a gives that our approaches with the shared execution schema, SPG, outperform the approaches without the shared execution paradigm, PPG, PRG, and the baseline algorithm, LPG and LRG in terms of query anonymization time. Since our shared execution schema dynamically finds a shared set of cloaked segments for a group of queries without any limitation on their locations, it generated a better anonymization cost for our shared execution algorithm. However, Due to it that our non-shared execution algorithms would first process a best query processing cost cloaked segment, our two algorithms would need more anonymization time compared to LPG and LRG, respectively. In general, the location anonymization time of all the approaches improves as there are more users in the system. This is due to the fact that smaller cloaked areas are generated to satisfy the same required K-anonymity level when the number of users increases.

Figure 1b gives that our approaches perform better than LPG and LRG in terms of Number of Open Nodes. Since LPG and LRG take into account the cost and quality of road network segments a peer value, the cost of LPG and LRG might not be the best. While our approaches would first deal with the query processing cost, and thus, our approaches give better query processing cost than LPG and LRG.

Figures 1c and 1d depict the resilience of all the approaches to the center-of-cloaked-area and replay privacy attacks, respectively. In terms of the center-of-cloaked-area privacy attack, all the approaches have very strong resilience. For the replay attack, it gives our approaches have a little stronger resilience than LPG

and LRG, this is because the re-run of our approach may have more changes while do the first best cost selections.

## 5.2 Number of Queries

Figure 2 gives the performance of all the approaches with respect to increasing the number of queries from 600 to 1,000.

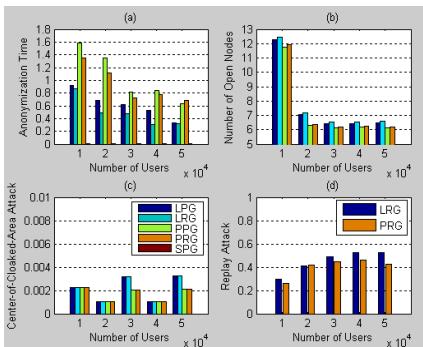
The anonymization time of the non-shared execution approaches, PPG, PRG, and the baseline approaches, LPG, LRG, is only slightly affected by the increase of the number of queries in the system, as depicted in Figures 2a. And our approaches provide better query processing cost than other approaches (Figures 2b).

Figures 2c and 2d shows that the resilience to the center-of-cloaked-area attack and the replay attack of the execution approaches, LPG, LRG, PPG, and PRG. In general, the success rate of center-of-cloaked-area attack decreases when the number of queries increases. The main reason is that the success number of attack is valid but the number of query users increases.

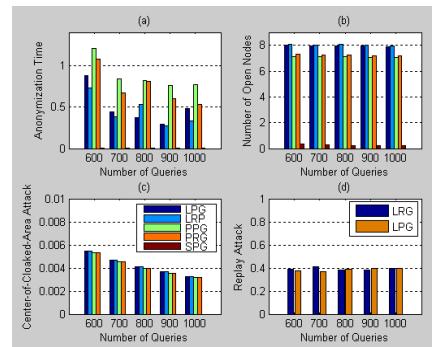
## 5.3 $k$ -Anonymity Privacy Requirements

Figures 3 depict the performance of all the approaches, as the user required  $k$ -anonymity level increases from 10 to 30.

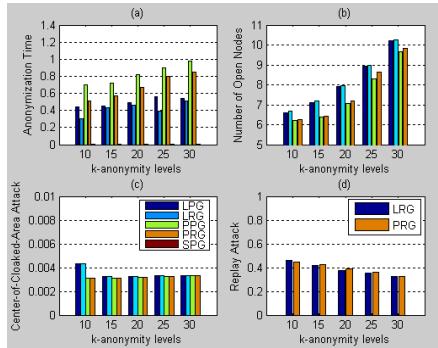
It is expected that when the location anonymization algorithm has to generate larger cloaked segment sets to satisfy the stricter privacy requirements, the location anonymization overhead of all the approaches increases (Figures 3a). Such larger cloaked segment sets lead to larger open nodes, so the query processing cost and query quality gets worse when  $k$  increases (Figure 3b). As is discussed before, larger cloaked segment sets lead to harder replay attack, is easily to know the success rate of replay attack decreases when  $k$  increases (Figure 3d). The success rate of center-of-cloaked-area attack may slightly change since the number of queries does not vary (Figure 3c).



**Fig. 1.** Number of Users



**Fig. 2.** Number of Queries



**Fig. 3.** k-anonymity levels

## 6 Conclusion

This paper designs a novel and effective anonymous model based on road network environment. In the paper, we propose a novel simple greedy algorithm that considers the trade-off between the query processing cost and the query quality, additionally guarantees the best of query processing cost. We evaluate our algorithm extensively through simulated experiments, which compared with the existing solution in terms of query anonymization time and number of open nodes, and the resilience of the two privacy model. The experimental results show that our location anonymization algorithms are efficient and scalable. In future work, we will continue to explore the various situations of the road network environment, and take into account the continuous queries and the processing of LBS.

## References

- [1] Cheng, R., Zhang, Y., Bertino, E., Prabhakar, S.: Preserving User Location Privacy in Mobile Data Management Infrastructures. In: Danezis, G., Golle, P. (eds.) PET 2006. LNCS, vol. 4258, pp. 393–412. Springer, Heidelberg (2006)
- [2] Chow, C.Y., Mokbel, M.F., He, T.: A Privacy-Preserving Location Monitoring System for Wireless Sensor Networks. IEEE Transactions on Mobile Computing, TMC (to appear)
- [3] Chow, C.Y., Mokbel, M.F., Liu, X.: Peer-to-Peer Spatial Cloaking for Anonymous Location-based Services. GeoInformatica (to appear)
- [4] Kalnis, P., Ghinita, G., Mouratidis, K., Papadias, D.: Preventing Location-Based Identity Inference in Anonymous Spatial Queries. IEEE Transactions on Knowledge and Data Engineering, TKDE 19(12), 719–733 (2007)
- [5] Sweeney, L.: k-anonymity: A Model for Protecting Privacy. International Journal of Uncertainty, Fuzziness and Knowledge -Based Systems, IJUFKS 10(5), 557–570 (2002)
- [6] Chow, C.-Y., Mokbel, M.F., Bao, J., Liu, X.: Query-Aware Location Anonymization for Road Networks. GeoInformatica (2011) (to appear)
- [7] Wang, T., Liu, L.: Privacy-aware mobile services over road networks. PVLDB 2(1), 1042–1053 (2009)
- [8] Papadias, D., Zhang, J., Mamoulis, N., Tao, Y.: Query processing in spatial network databases. In: VLDB, pp. 802–813 (2003)
- [9] Zhang, C., Huang, Y.: Cloaking Locations for Anonymous Location based Services: A Hybrid Approach. GeoInformatica 13, 159–182 (2009)

# A Wireless Measurement System (M3D) for Three-Dimensional Gait Analysis System

Tao Liu<sup>1</sup>, Yoshio Inoue<sup>1</sup>, Kyoko Shibata<sup>1</sup>, and Kouzou Shiojima<sup>2</sup>

<sup>1</sup> Dep. of Intelligent Mechanical Systems Engineering  
Kochi University of Technology, Kami, Japan  
liu.tao@kochi-tech.ac.jp

<sup>2</sup> Tec Gihan Co., Ltd  
Nishinohata, Okubo-Cho, Uji-City, 611-0033, Kyoto, Japan  
k.shiojima@tecgihan.co.jp

**Abstract.** In order to implement three-dimensional (3D) gait analysis, a complete human kinematic analysis using inertial sensors is not enough, and a mobile force plate system to measure ground reaction force (GRF) during successive gaits is necessary for inverse human dynamics analysis. In this paper, a complete 3D gait analysis based on a wireless measurement system is proposed. The measurement system named as M3D was developed by integrating a mobile force plate, 3D motion analysis units based on MEMS sensors and a wireless data logger. A stick-chain model was built to visually analyze 3D human gait and joint trajectories, and triaxial joint moments during gait can be calculated.

**Keywords:** Gait, MEMS, Human dynamics, Joint moments.

## 1 Introduction

In order to implement three-dimensional (3D) gait analysis, a complete human kinematic analysis using inertial sensors is not enough, and a mobile force plate system to measure ground reaction force (GRF) during successive gaits is necessary for inverse human dynamics analysis. By mounting multi-axial force sensors beneath a special shoe, some instrumented shoes have been developed for ambulatory measurements of triaxial GRF in a variety of non-laboratory environments. To analyze dynamics gait and joint loads, 3D inertial sensor modules have been integrated into wearable force plates. Liedtke et al. proposed a combination sensor system including six degrees of freedom force and moment sensors and miniature inertial sensors provided by Xsens Motion Technologies to estimate joint moments and powers of the ankle [1]. In our past research, a thin and light force plate based on triaxial sensors and inertial sensors was also proposed to analyze continuous gaits by measuring triaxial GRF and foot orientations [2], [3]. We are presently concentrating on the development of some wearable sensors to measure human GRF and segment orientations during gait. If 3D orientations of all the leg segments are integrated with the measured triaxial GRF, an inverse dynamic method can be used to implement joint dynamics analysis of lower limb.

In this paper, a complete 3D gait analysis based on a wireless sensor system is proposed. The sensor system named as M3D was developed by integrating a mobile force plate, 3D motion analysis units based on MEMS sensors and a wireless data logger. A stick-chain model was built to visually analyze 3D human gait and joint trajectories.

## 2 Methods

As shown in Fig. 1, a small motion sensor unit (weight: 20g, size: 35×50×15mm<sup>3</sup>) was designed using a triaxial accelerometer, three uniaxial gyroscopes and a triaxial magnetic sensor and micro-computer system, which were provided by Tec Gihan Co. Japan. The sensor unit can communicate with a data transfer or personal computer by a RS-485 serial communication port. The inertial and magnetic sensors' specification parameters are given in Table 1. Nine channels sensor signals (triaxial accelerations, triaxial angular rates, and triaxial magnetic intensities) are provided after a 16-bit A/D conversion.



**Fig. 1.** Prototype of a motion sensor unit

**Table 1.** Main specifications of inertial and magnetic sensors

	Accelerometer	Gyroscope	Magnetic sensor
Capacity	± 8g	± 1200°/s	± 70µT
Nonlinearity	±0.5%	± 1%	± 0.1%
Response	2kHz	140Hz	10kHz

A Kalman-based fusion algorithm has been applied to process signals of the triaxial accelerometer and triaxial magnetic sensor by incorporating excellent dynamics of gyroscope and stable drift-free performance of the accelerometer and magnetic sensors. 3D orientations of the sensor units when mounted on human body segments can be calculated using the filtered signals. In order to remove the effects from motion accelerations and measurement errors on accelerometer measuring the gravity acceleration, we design an EKF algorithm using (1) and (2), and the state variables and measurement vector are given. Since the estimated variables are derived from the hybrid of accelerometer and gyroscope, it incorporates excellent dynamics of the gyroscope measurement and stable drift-free performance of gravity acceleration using the accelerometer.

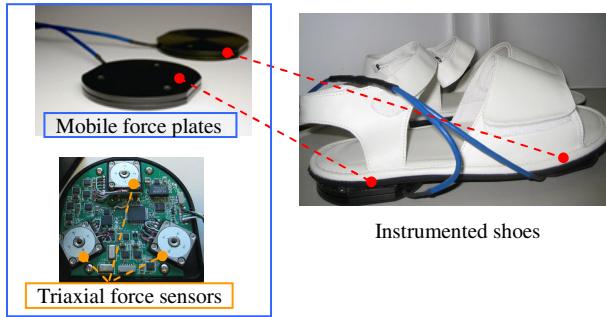
$$\begin{bmatrix} A_g^X(k) \\ A_g^Y(k) \\ A_g^Z(k) \end{bmatrix} = \begin{bmatrix} 1 & T \cdot \omega^Z(k-1) & -T \cdot \omega^Y(k-1) \\ -T \cdot \omega^Z(k-1) & 1 & T \cdot \omega^X(k-1) \\ T \cdot \omega^Y(k-1) & -T \cdot \omega^X(k-1) & 1 \end{bmatrix} \begin{bmatrix} A_g^X(k-1) \\ A_g^Y(k-1) \\ A_g^Z(k-1) \end{bmatrix} \quad (1)$$

$$\begin{bmatrix} Z^X(k) \\ Z^Y(k) \\ Z^Z(k) \end{bmatrix} = \begin{bmatrix} A_g^X(k) \\ A_g^Y(k) \\ A_g^Z(k) \end{bmatrix} \quad (2)$$

$k=1, 2, 3\dots$

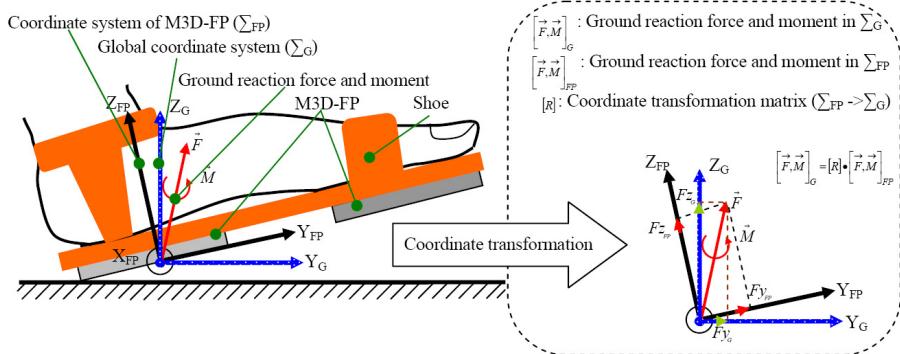
where  $[A]$  is the state vector and  $[Z]$  denotes the measurement acceleration vector directly equal to the accelerometer measurement vector;  $[\omega]$  is a 3D angular velocity vector obtained from the gyroscope measurements.

Small triaxial force sensors (USL06-H5-500N) provided by Tec Gihan Co. Japan can only detect the three-directional force induced on a small circular plate ( $\Phi$  6 mm), so it is difficult to apply directly to the measurement of the GRF distributed under feet. As shown in the right photos of Fig. 2, a mobile force plate (weight: 110g, size: 82×88×9mm<sup>3</sup>) to measure triaxial force and triaxial moment was developed using the three small triaxial force sensors, in which two aluminum plates were used as top and bottom plates to accurately fix the three sensors and signal processing circuits. A detailed description of the method to extract the triaxial GRF can be found in our previous publications [2]. In this research, range of force measurement of the developed force plate in instrumented shoes for the vertical direction and two horizontal directions is 1000N and 500N, respectively. The maximum torque measured by the force plate is 30Nm for all directions.



**Fig. 2.** Prototype of an instrumented shoe system with two mobile force plates mounted under the heel and forefoot

In order to implement ambulatory GRF measurements when the force plates move with feet, a 3D motion sensor unit based on MEMS sensors to measure 3D orientations of the mobile force plate was added inside the force plate. The motion sensor unit can measure triaxial accelerations, angular velocities and magnetic vector, and data from the motion sensors can be combined with force sensors' data for a dynamic GRF measurement. As shown in Fig. 3, coordinate transformation from Coordinate system of M3D-FP ( $\Sigma_{FP}$ ) to Global coordinate system ( $\Sigma_G$ ) is implemented by the measures of the force plate system.

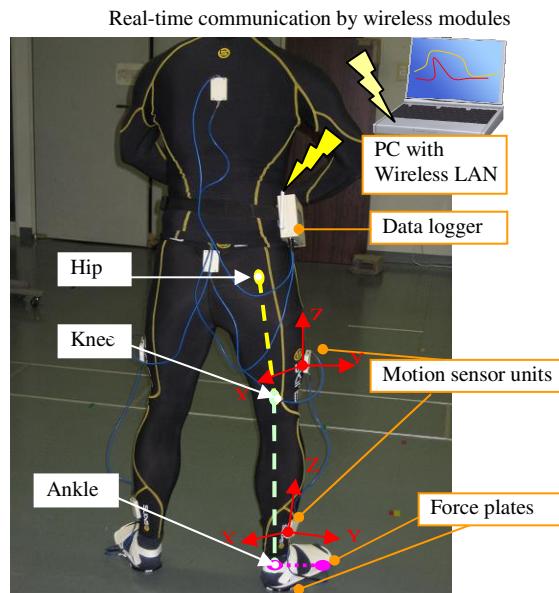


**Fig. 3.** Coordinate transformation from Coordinate system of M3D-FP ( $\Sigma_{FP}$ ) to Global coordinate system ( $\Sigma_G$ )

### 3 Experiments

#### 3.1 Human Motion and Force Measurements Using M3D

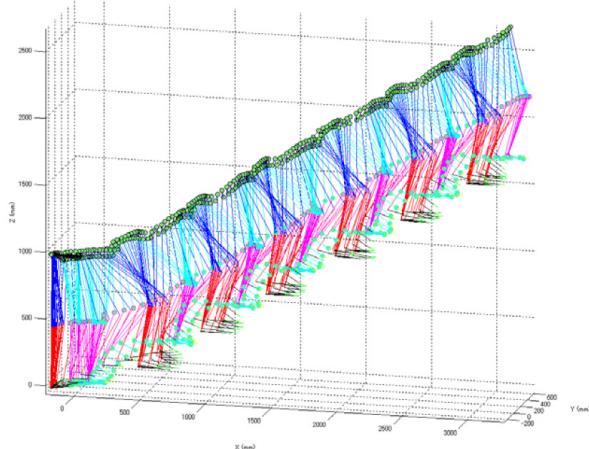
As shown in Fig. 4, we used the 3D motion sensor units to measure orientations of the shank and the thigh of two legs, and the instrumented shoes were worn by subjects to measure GRF, and foot segments' orientations.



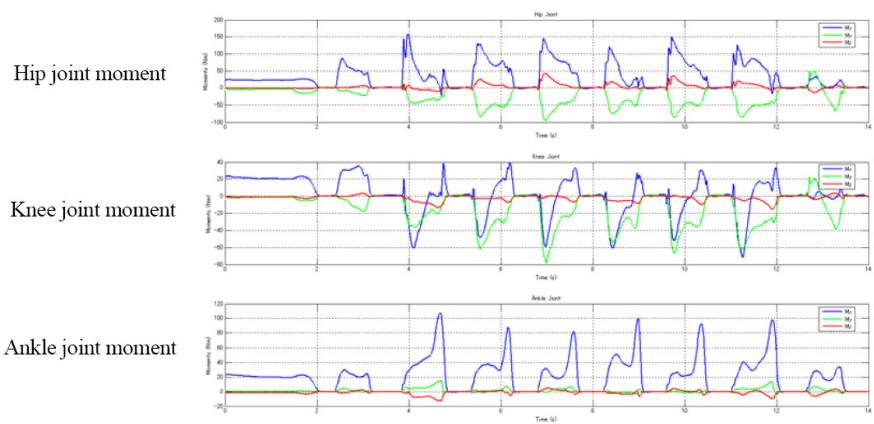
**Fig. 4.** Measurement system (M3D) for 3D gait analysis

### 3.2 Joint Moments

3D lower limb gait posture was calculated with the segment orientations, detections of gait phase cycle, and lengths of the leg segments. As shown in Fig. 5 and Fig. 6, a stick-chain model is used to visually analyze the lower limb postures, and a group of representative results of joint moments in stair climbing trials are given.



**Fig. 5.** Stick-link results of stair climbing



**Fig. 6.** Joint moments during stair climbing

## 4 Conclusions

As an alternative tool of the traditional gait analysis system based on high-speed cameras and stationary force plates, a wireless sensor system was developed to obtain

3D motion and force data on successive gait in various walking environments, and a quantitative gait analysis method based the ambulatory measurements is proposed to implement human lower limb kinematic and kinetic analysis. The visualized body segment orientation and 3D joint moment data should be helpful to medical doctors in monitoring and evaluating patient recovery status.

## References

1. Liedtke, C., Fokkenrood, S.A.W., Menger, J.T., van der Kooij, H., Veltink, P.H.: Evaluation of instrumented shoes for ambulatory assessment of ground reaction forces. *Gait and Posture* 26(1), 39–47 (2007)
2. Liu, T., Inoue, Y., Shibata, K.: A wearable force plate system for the continuous measurement of triaxial ground reaction force in biomechanical applications. *Measurement Science and Technology* 20(8), no. 085804 (2010)
3. Liu, T., Inoue, Y., Shibata, K.: Development of a Wearable Sensor System for Quantitative Gait Analysis. *Measurement* 42(7), 978–988 (2009)

# Approaches and Controllers to Solving the Contention Problem for Packet Switching Networks: A Survey

Fouad Kharroubi<sup>1</sup>, Lin Chen<sup>1</sup>, and Jianjun Yu<sup>2</sup>

<sup>1</sup> School of Information Science and Engineering, Hunan University, Changsha, China

<sup>2</sup> NEC Laboratories America, Princeton, New Jersey, USA

fouad.kharroubi@gmail.com, liliuchen12@vip.163.com,  
jianjun@ece.gatech.edu

**Abstract.** Optical packet switching (OPS) is a promising technology for future networks. However, optical packet contention is a major problem in an OPS network. Resolution and avoidance are two schemes that can deal with the contention problem. A resolution scheme, as a reactive approach, resolves collisions, while an avoidance scheme, as a proactive approach, tries to reduce the number of potential collision events. Therefore, many contention controllers using neural networks have been proposed to control the output contention problem within a learning approach. In this article, we survey the contention resolution and avoidance schemes proposed for OPS networks. We also review some contention controller propositions using neural network techniques for solving the output contention problem in OPS.

**Keywords:** Optical Packet Switching (OPS), contention avoidance, contention resolution, neural networks, contention controllers.

## 1 Introduction

The Internet has already reached an enormous complexity, and it is still rapidly growing. A wide variety of new services have emerged. Real time and multimedia applications, such as Internet telephony, videoconferencing, video-on-demand and many others, make high demands on future networking solutions, such as increasing capacity and appropriate quality of service (QoS) provisioning [1]. Meanwhile, wavelength-division multiplexing (WDM) technology [2] has been widely deployed as an optics solution to meet the demand for high-bandwidth networking, providing an attractive platform to exploit the bandwidth potential of fiber links and it is clear that for the next generation of the optical internet [3], the focus is now moving from circuit switched networks, which occupy a wavelength continuously, regardless of the demand at that time, towards optical packet/burst switching (OPS/OBS) by only occupying a wavelength when data is to be transmitted [1], [3]. We are striving for a more efficient utilization of bandwidth in optical fibers. Hence an important QoS

attribute is the packet loss ratio due to the contention problem associated with optical networks.

Indeed, contention avoidance and resolution schemes are key determinants of packet loss performance in any packet switching paradigm [2], [8]. Contention arises in an OPS router if two or more packets compete for the same output fiber on the same wavelength at the same time. Various techniques have been examined in literature. Some of these proposed solutions follow a reactive approach [5], [6] to addressing the contention problem such as wavelength conversion, optical buffering, deflection routing, and retransmission [4], [2]. They focus on resolving the collisions only after they occur. Other proposed solutions take a proactive approach, [2], [4], [7] aiming at avoiding/minimizing the occurrence of contentions inside the OPS networks.

The objective of this article is to survey the different contention avoidance and resolution schemes existing in literature and we also aim to shed more light on the neural-network controllers as a learning approach to solving the output contention problem in packet switching networks. This article is organized as follows: existing contention resolution and avoidance schemes are surveyed in Section 2; and Section 3 reviews some neural network controllers as a learning approach to solving the output contention problem in packet switching networks; and our conclusions are discussed in Section 4.

## 2 Approaches to Solving the Contention Problem

Since in OPS optical packets from different source nodes arrive at a core switch without any coordination among transmitters, contention may occur when two or more packets want to use the same output link of an OPS core switch at the same time. If no mechanism is used to address the contention problem, all contending packets except the one that has arrived first, have to be dropped, which can result in an unacceptable performance of the node and the whole network. To deal with this contention problem, resolution and avoidance are two common approaches that will be surveyed briefly in this section. More details can be found in [8].

### 2.1 Contention Resolution Schemes: A Reactive Approach

Most of the approaches used to control the packet contentions are based on reactive methods. In other words, they try to resolve the problem only after the contentions occur. Therefore, these techniques are usually referred to in literature as contention resolution schemes. A number of hardware-based and software-based schemes have been proposed to reduce the number of collision events in an OPS network [8].

#### **Wavelength Conversion**

Consider a situation in which an optical packet arriving on a given wavelength channel is blocked because the channel with the same wavelength on the appropriate output link is busy. In this situation, an OPS core switch can use WCs (Wavelength

Converter) to resolve the blockages of optical packets by transmitting a contending optical packet on another wavelength. Obviously, avoiding blockages reduces the PLR [8]. The wavelength converters, as hardware-based contention resolution schemes, can operate with a full degree (i.e., they can convert any incoming wavelength to a fixed desired wavelength) or with a limited range (i.e., they can convert one or several pre-determined incoming wavelengths to a fixed desired wavelength) [10]. Different wavelength conversion architectures have been proposed in literature. Table 1 displays a refinement of the well-used classification of the wavelength conversion architectures [8] [10].

**Table 1.** Classification of the Wavelength Conversion Architectures

Full wavelength conversion [13], [14]
Shared wavelength conversion
Shared-per-node wavelength conversion [13], [17]
Shared-per-link wavelength conversion
Shared-per-input-fiber [20], [21], [22]
Shared-per-output-fiber [13], [23]
Limited-range wavelength conversion [15], [16]

### FDL Buffering

In conventional packet-switched networks, buffering is the main approach for circumventing the contention problem. Applying the same technique to the optical domain is, however, quite a challenging task, since there is no equivalent to random access memory in the optical domain. Instead, optical delay loops can be used to delay packets during times of contention. A delay loop is realized through a simple piece of fiber optic and, therefore, the provided delay is fixed and inflexible in the sense that once a packet is forwarded to the delay loop, it cannot be retrieved before it leaves the loop from the other end [10], [12].

A fiber delay line (FDL) buffer can be built by employing several basic delay lines of different lengths in a parallel structure. Depending on the type of the basic delay lines applied, FDL buffers can be classified into two major categories, namely, fixed-delay and variable-delay buffers [12]. In a fixed-delay buffer [24], [25], each basic delay line provides only a fixed. Alternatively, in a variable-delay buffer [26], [27], each basic delay line is a multistage structure. However, it should be noted that some authors, such as in [10], consider that using multistage, multilength optical delay lines, or using a large number of optical delay lines, are complex optical buffer arrangements that introduce an extra hardware requirement and an additional physical impairment to the optical signal.

### Deflection Routing

This technique is considered to be a software-based contention resolution scheme that uses the idea of routing to resolve contention and that imposes no additional hardware cost on the network [8]. It relies on a neighboring node to route the packet when contention occurs [10]. In other words, the deflection routing consists of the use of all the links across the network as shared buffers. Specifically, in a core switching node

that implements deflection routing, those packets that lose the competition on accessing the appropriate channels (links) are forwarded over other links to another core switch in the network [12].

This technique depends on the topology of the core switch. Deflection routing is cheap and simple but if a deflected packet takes a longer path to reach its destination, this leads to a higher end-to-end delay as well. In addition, the packets are likely to arrive out of sequence at the destination [28]. The deflected packets can also cause network congestion, especially at high traffic loads [29]. Interested readers will find in [8], [30], [31], [32] more details about deflection routing on regular network topologies.

### **Retransmission in the Optical Domain**

Retransmission, as software-based contention resolution schemes, is another technique to recover the lost traffic in which each ingress switch keeps a copy of the transmitted traffic in its electronic buffer and retransmits whenever required [8]. There are two retransmission techniques in the optical domain. The conventional retransmission technique (re)transmits optical packets until there is a successful transmission. A prioritized retransmission technique is also proposed and analyzed for slotted OPS. In this technique, each packet that is transmitted by an ingress switch toward an egress switch carries a priority field that gives a better chance to the packets in longer-hop connections to pass through the OPS network. More details can be found in [8], [33], [34], [35], [36].

### **Hybrid Schemes**

Although any of the three basic contention resolution schemes in optical networks (i.e., wavelength conversion, FDL buffering and deflection routing) can help reduce contentions and improve performance, relying merely on one scheme might not suffice to achieve an acceptable performance or might be, in some situations, too costly. Thus, many authors have proposed many hybrid schemes by incorporating two or all the basic schemes [12]. Among all possible combinations, there are two hybrid schemes that are considered the most promising, namely, wavelength conversion with FDL buffering and wavelength conversion with deflection routing [37]. Between these two hybrid schemes, performance comparison studies have revealed that the option with FDL buffers outperforms the option with deflection routing [10].

## **2.2 Contention Avoidance Schemes: A Proactive Approach**

A number of hardware-based and software-based schemes have been proposed as a proactive approach to avoid/minimize the occurrence of contentions inside the OPS networks. Therefore, these techniques are usually referred to in literature as the contention avoidance schemes. We, then, present only the most often used approaches dealing with this problem.

### **Designing a Multi-fiber Architecture**

By adopting this technique as a hardware-based contention avoidance scheme, the researchers aim to converge toward results in a lower number of contentions and a lower number of optical packet drops than a single-fiber network. Thus, in this

technique each connection link in an OPS network, between any pair of edge-to-core switches or any core-to-core switches, contains  $f$  fibers with a small number of wavelengths per fiber [8]. On the other side the goal of this technique is achieved by increasing the hardware components and, therefore, the network cost. More details can be found in [38], [39], [40].

### **Using Additional Wavelengths to Carry the Same Traffic**

Another technique to lower traffic load in order to reduce PLR without affecting the traffic load is to use additional wavelengths to transmit the same traffic on each connection link. The use of additional wavelengths in a fiber provides more bandwidth to carry the same traffic [8]. However, this requires more transceivers at ingress and egress switches and more devices to handle a higher number of wavelengths on a fiber [39], [41], [42].

### **Controlled Traffic Transmission**

This proactive approach aims to avoid the contention problem consisting of several of the techniques in which each ingress switch sends its traffic to an OPS network in a controlled manner. Among these techniques are those called coordinated packet transmission, in which each ingress switch sends its traffic on a given wavelength in such a way that the variance of the number of optical packets arriving at the given core switch within any time-slot is minimized. However this scheme is much more effective at lower traffic loads, and the effectiveness of this technique is reduced by increasing the traffic load [8], [43].

Symmetric traffic transmission [8], [39], [43] is another technique in which each ingress switch connected to a given core switch must send its traffic in a symmetric manner (i.e., with the same probability) to different output links of the given core switch. Otherwise, when traffic distribution to the output links of the given core switch is asymmetric, the PLR is high. Load-balanced traffic transmission [8], [39], [43] is also another way used so that the traffic on each connection link between either an edge switch to a core switch or a core switch to a core switch can be transmitted in a balanced manner so that no wavelength is overwhelmed.

The last technique we will include in this section is called monitoring core switch traffic; in this technique, an ingress switch connected to a given core switch keeps client packets in its electronic transmission buffers. On the other hand, a scheduler located at the given core switch monitors the status of the wavelengths at its output ports. When there is no transit packet occupying a desired wavelength in an output port, the scheduler informs the ingress switch to send its packet toward the given core switch [8]. By this monitoring of optical packets and not sending those packets that are likely to cause collision in the given core switch, the ingress switch reduces the PLR at the given core switch. Note that in this technique the given core switch needs a complex controller.

## **3 Controllers to Solving the Contention Problem**

In the previous sections, we have dealt with how the contention problem in OPS can be solved by choosing either a reactive or proactive approach. A number of hardware-based

and software-based schemes have been proposed to avoid/minimize the occurrence of contentions inside the OPS networks. In this section, we shall review available literature, which use the learning approach, to implement the contention controller based on the neural- networks to solve the output contention in packet switching networks.

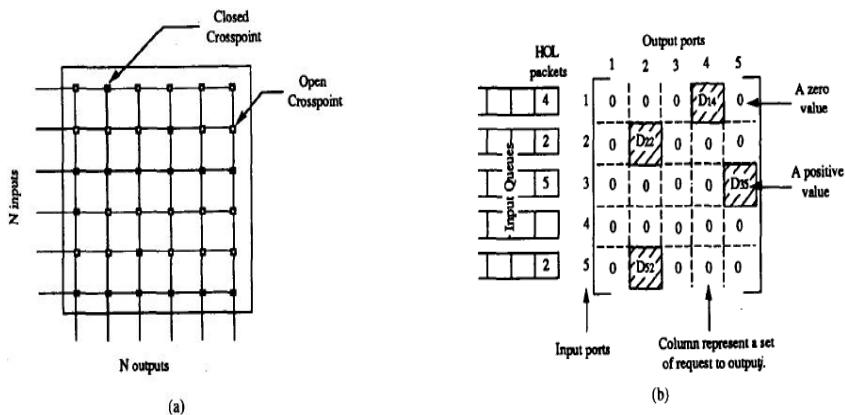
### **3.1 Neural Networks for Switching**

Neural networks are an extraction of computational ideas from the field of neurobiology. They are parallel processors that consist of many extremely simple highly interconnected processing elements [48]. The neural network derives its computing power from both its many processors and the high connectivity that allows information to flow rapidly through the network. This massive parallelism implies that neural networks can greatly increase the speed of the computation [48]. Neural networks are seen also as a class of system that has many simple processors-neurons, which are highly interconnected. The function of each neuron is simple, and the behavior is determined predominately by the set of interconnections [44]. Thus, a neural network is a special form of parallel computer. Although a major impetus for using neural networks is that they may be able to “learn” the solution to the problem that they are to solve, another, perhaps even stronger, impetus is that they provide a framework for designing massively parallel machines [44]. The highly interconnected architecture of switching networks suggests similarities to neural networks [44]. Next, we will present some approaches to solving the contention problem using neural networks.

### **3.2 Neural Network for Solving the Contention Problem**

Output contention or output blocking is an inherent feature of packet switches due to the unscheduled arrival process of various packets. In this section we will present an interesting approach that has been proposed in [45] to resolve the output contention in a packet switch with a synchronous switching mode, i.e., the connection requests from all of the head-of-line (HOL) packets are simultaneously presented to the switch. Indeed, output contention occurs when there are more than one (up to N) HOL packets from the input ports simultaneously requesting to be switched to the same output port. Only C of these contending packets can be routed to the same output port. The Approach [45] taken in resolving the output-contention problem and selecting up to C packets for each output port depends on the service selection policy (called also the service discipline) employed by the packet switch. Typical service selection policies, cited in [45], are the random selection policy, the FIFO selection policy, the longest queue selection policy and the customer priority selection policy.

This approach to resolving the output contention and implementing the service-selection policy was described in [45] by referring to Fig. 1, which shows how a request for a packet transmission through an  $N \times N$  crossbar can be mapped onto an  $N \times N$  input request matrix. Each input and output connection of the nonblocking optoelectronic WDM packet switches can be represented by an imaginary crosspoint.



**Fig. 1.** (After Ref. [45]): (a) a crossbar switch used to illustrate the input-output connections in the WDM packet switch. (b) Input request matrix describing the state of the packets in the HOL positions.

The  $N \times N$  input request matrix describes the input request state of the HOL packets. The rows and columns of the input request matrix correspond to the input and output ports, respectively. If there is a packet from input port  $i$  destined to output port  $j$ , then cell  $D_{ij}$  in the input request matrix is assigned a positive number, otherwise  $D_{ij} = 0$ . The value of  $D_{ij}$  will depend on the type of service selection policies used. More information about this mathematical model can be found in [45].

### The K-Winner Neural Network for Solving the Contention Problem

The key to determining suitable neural network architecture is to match the network topology to the contention problem as closely as possible [44], [45]. Since the problem of finding the nonblocking connections configuration is a constraint issue, i.e., selecting up to  $C$  largest value cell matrix from each column in the HOL input request matrix, in [45] they proposed to use the K-winner network. The K-winner network falls in the category of self-organizing learning algorithms, and it has the property that, given  $N$  neurons, all initially off and with the initial state, only  $K$  neurons with larger external inputs will turn on. With each HOL input request matrix column represented by a K-winner network and each of the  $N$  neurons corresponding to the  $N$  rows of cell, the nonblocking connection configuration of the switch can then be easily solved. Therefore, there is a neuron for every crosspoint in the  $N \times N$  crossbar shown in Fig. 1. Interested readers will find in [45] more details about how to implement the contention controller based on the K-winner network and its interface with the input and output port controllers for the Neuro-Star packet switch.

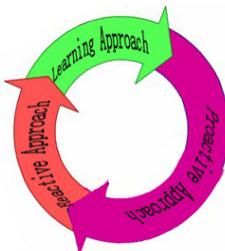
### Other Neural Network for Solving the Contention Problem

Other neural networks have been considered for solving the contention problem in packet switching. Indeed, Troudet et al. [46] and Marrakchi et al. [47] have proposed the use of a Hopfield neural network for real time control of an input queued crossbar switch for switching packets at maximum throughput. Brown et al. [48] has proposed

a neural network based on multiple overlapping winner-take-all circuits to compute a nonblocking switching configuration for an input queued Banyan switch with a performance within a factor of two of the nonblocking switch. A. Badi et al. [49] have proposed the same idea proposed in [45] but use another neural network technique called Order Statics Filter (OSF) for designing a contention controller with a speed up factor to achieve a real-time computation of a non-blocking switching high-speed high-capacity packet switch without packet loss.

## 4 Conclusion

We have surveyed the contention resolution, as a reactive approach, and the avoidance schemes, as a proactive approach, to overcome the contention problem in OPS networks. For each technique, we have briefly explained its basic principle and have given many references to where the interested readers can find more details. Some neural network controllers as a learning approach to solving the output contention problem in packet switching networks also have been investigated. We think that ever that one of the most effective solutions that could be proposed for solving the output contention problem is to combine the contention resolution schemes with the contention avoidance schemes at the same time by taking a three-dimensional cyclic approach : proactive-reactive and learning (Fig. 2). This new proposal will be explained and detailed in our next communication.



**Fig. 2.** Three-dimensional cyclic approach positioning

## References

1. Heegaard, P.E., Sandmann, W.: Evaluating Differentiated Quality of Service Parameters in Optical Packet Switching. In: Koucheryavy, Y., Harju, J., Sayenko, A. (eds.) NEW2AN 2007. LNCS, vol. 4712, pp. 162–174. Springer, Heidelberg (2007)
2. Fei, X., Ben Yoo, S.J.: High-Capacity Multiservice Optical Label Switching For the Next-Generation Internet. IEEE Communication Magazine 42(5) (2004)
3. Cheyns, J.C., Develder, E., Van, B., Ackaert, A., Pickavet, M., Demeester, P.: Routing in an AWG-based Optical Packet Switch. Photonic Network Communications 5(1), 69–80 (2003), doi:10.1023/A:1021005913665

4. Lévesque, M., Elbiaze, H., Wael, H.F.A.: Adaptive Threshold-based Decision for Efficient Hybrid Deflection and Retransmission Scheme in OBS Networks. In: 13th Conference on Optical Network Design and Modeling (ONDM 2009), Braunschweig, Germany, February 18-20 (2009)
5. Rostami, A., Chakraborty, S.S.: On Performance of Optical Buffers with Specific Number of Circulations. *IEEE Photonics Technology Letters* 17(7), 1570–1572 (2005)
6. Rostami, A., Wolisz, A.: Modeling and Synthesis of Traffic in Optical Burst- Switched Networks. *IEEE/OSA Journal of Lightwave Technology (JLT)* 25(10), 2942–2952 (2007)
7. Rostami, A., Wolisz, A., Feldmann, A.: Traffic analysis in optical burst switching networks: a trace-based case study. *European Transactions on Telecommunications (ETT)* 20(7), 633–649 (2009) (invited paper)
8. Ghaffar, A., Oliver, W., Yang, W.: Contention Avoidance and Resolution Schemes in Bufferless All-Optical Packet-Switched Networks: A Survey. *IEEE Communications Surveys & Tutorials* 10(4) (Fourth Quarter, 2008)
9. Lévesque, M., Elbiaze, H.: Graphical Probabilistic Routing Model for OBS Networks with Realistic Traffic Scenario, Department of Computer Science, Quebec University (QC) Canada. In: *IEEE GLOBECOM Proceedings* (2009)
10. Yao, S., Biswanath, M., Ben Yoo, S.J., Sudhir, D.: A Unified Study of Contention-Resolution Schemes in Optical Packet-Switched Networks. *Journal of Lightwave Technology* 21(3) (March 2003)
11. Rarnaswami, R., Sivarajan, K.: Optical Networks: A practical perspective. Morgan Kaufmann (2001)
12. Rostami, A.: A dissertation about Traffic Shaping for Contention Control in OBS Networks. Presented to the Academic Faculty in Berlin (2010)
13. Eramo, V., Listanti, M., Pacifici, P.: A Comparison Study on the Number of Wavelength Converters Needed in Synchronous and Asynchronous All-Optical Switching Architectures. *J. Lightwave Tech.* 21(2) (2003)
14. Øverby, H.: Performance Modeling of Synchronous Bufferless OPS Networks. In: Proc. Int'l. Conf. Transparent Optical Networks, Wroclaw, Poland (July 2004)
15. Eramo, V., Listanti, M., Donato, M.D.: Performance Evaluation of a Bufferless Optical Packet Switch with Limited-Range Wavelength Converters. *IEEE Photonic Technology Letters* 16(2), 644–645 (2004)
16. Almeida, R.C., Martins-Filho, J.F.: Limited-Range Wavelength Conversion Modeling for Asynchronous Optical Packet-Switched Networks. In: SBMO/IEEE MTT-S Int'l. Conf. Microwave and Optoelectronics, Kharkov, Ukraine (2005)
17. Eramo, V.: Performance of Scheduling Algorithms in Optical Packet Switches Equipped with Limited-Range Wavelength Converters. *J. Optical Networking* 4(12), 856–869 (2005)
18. Lee, K.C., Li, V.O.K.: Optimization of a WDM Optical Packet Switch with Wavelength Converters. In: Proc. IEEE INFOCOM, Boston, Massachusetts (April 1995)
19. Eramo, V., Listanti, M., Spaziani, M.: Dimensioning Models in Optical Packet Switches Equipped with Shared Limited-Range Wavelength Converters. In: IEEE Globecom 2004, Dallas, USA (December 2004)
20. Danielsen, S.L., Hansen, P.B., Stubkjaer, K.E.: Wavelength Conversion in Optical Packet Switching. *J. Lightwave Tech.* 16(12) (December 1998)
21. Eramo, V., Listanti, M., Tarola, M.: Advantages of Input Wavelength Conversion in Optical Packet Switches. In: Proc. IEEE Globecom 2003, San Francisco, CA, USA (December 2003)
22. Callegati, F., Corazza, G., Raffaelli, C.: Exploitation of DWDM for optical Packet Switching with Quality of Service Guarantees. *IEEE JSAC* 20(1), 190–201 (2002)

23. Dogan, K., Akar, N.: A Performance Study of Limited-Range Partial Wavelength Conversion for Asynchronous Optical Packet/Burst Switching. In: Proc. IEEE ICC, Istanbul, Turkey (2006)
24. Tancevski, L., Yegnanarayanan, S., Castanon, G., Tamil, L., Masetti, F., Mc-Dermott, T.: Optical routing of asynchronous, variable length packets. *IEEE Journal on Selected Areas in Communications* 18(10), 2084–2093 (2000)
25. Chia, M.C., Hunter, D.K., Andonovic, I., Ball, P., Wright, I., Ferguson, S.P., Guild, K.M., O’Mahony, M.J.: Packet loss and delay performance of feedback and feed-forward arrayed-waveguide gratings-based optical packet switches with WDM inputs-outputs. *Journal of Lightwave Technology* 19(9) (2001)
26. Yoo, M., Qiao, C., Dixit, S.: QoS performance of optical burst switching in IP-over-WDM networks. *IEEE Journal on Selected Areas in Communications* 18(10), 2062–2071 (2000)
27. Chlamtac, I., Fumagalli, A., Suh, C.J.: Multibuffer delay line architectures for efficient contention resolution in optical switching nodes. *IEEE Transactions on Communications* 48(12), 2089–2098 (2000)
28. Xu, L., Perros, H., Rouskas, G.: Techniques for Optical Packet Switching and Optical Burst Switching. *IEEE Commun. Mag.*, 136–142 (January 2001)
29. El-Bawab, T.S., Shin, J.: Optical Packet Switching in Core Networks: Between Vision and Reality. *IEEE Commun. Mag.* 40(9), 60–65 (2002)
30. Yao, S., Mukherjee, B., Dixit, A.: Advances in Photonic Packet Switching: an Overview. *IEEE Commun. Mag.* 38, 84–94 (2000)
31. Fayoumi, A.G., Jayasumana, A., Sauer, J.: Performance of Multihop Networks Using Optical Buffering and Deflection Routing. In: Proc. IEEE Conf. Local Computer Networks (LCN 2000), Tampa, Florida, USA (November 2000)
32. Pattavina, A.: Performance of Deflection Routing Algorithms in IP Optical Transport Networks. *Computer Networks* 50(2), 207–218 (2006)
33. Modiano, E.: Random Algorithms for Scheduling Multicast Traffic in WDM Broadcast-and-Select Networks. *IEEE/ACM Trans. on Networking* 7(3), 425–434 (1999)
34. Zhang, Q., et al.: Evaluation of Burst Retransmission in Optical Burst-Switched Networks. In: Proc. IEEE Broadnets 2005, Boston, USA (October 2005)
35. Rahbar, A.G.P., Yang, O.: Retransmission in Slotted Optical Networks. In: Proc. IEEE High Performance Switching and Routing (HPSR), Poznan, Poland, pp. 21–26 (June 2006)
36. Rahbar, A.G.P., Yang, O.: Prioritized Retransmission in Slotted All-Optical Packet-Switched Networks. In: High Performance Switching and Routing, Workshop (2006), doi:10.1109/HPSR.2006.1709675
37. Aracil, J., Callegati, F.: Enabling Optical Internet with Advanced Network Technologies. Springer (2009)
38. Li, Y., Xiao, G., Ghafouri-Shiraz, H.: On the Benefits of Multifiber Optical Packet Switch. *Microwave and Optical Technology Letter* 43(5), 376–378 (2004)
39. Rahbar, A.G.P., Yang, O.: Contention Avoidance in Slotted Optical Networks. In: Proc. Int’l. Conf. Opt. Commun. Systems and Networks, SPIE Photonics North, Toronto, Canada, vol. 5970 (September 2005)
40. Rahbar, A.G.P., Yang, O.: Fiber-Channel Tradeoff for Reducing Collisions in Slotted Single-Hop Optical Packet-Switched (OPS) Networks. *OSA J. Optical Networking* 6(7) (July 2007)
41. Gerstel, O., Raza, H.: Merits of Low-Density WDM Line Systems for Long-Haul Networks. *J. Lightwave Tech.* 21, 2470–2475 (2003)

42. Berry, R., Modiano, E.: Reducing Electronic Multiplexing Costs in SONET/WDM Rings with Dynamically Changing Traffic. *Journal on Selected Areas in Communications* 18(10) (October 2000)
43. Rahbar, A.G.P., Yang, O.: Reducing Loss Rate in Slotted Optical Networks: A Lower Bound Analysis. In: Proc. IEEE ICC 2006, Istanbul, Turkey (June 2006)
44. Timothy Brown, X.: Neural Networks for Switching. *IEEE Communications Magazine* (November 1989)
45. Binh, L.N., Chong, H.C.: A Neural-Network Contention Controller for Packet Switching Networks. *IEEE Transactions on Neural Works* 6(6) (November 1995)
46. Troudet, T.P., Walters, S.M.: Neural Network Architecture for Crossbar Switch Control. *IEEE Transactions on Circuits and Systems* 38(1) (January 1991)
47. Marrakchi, A., Troudet, T.: A Neural Net Arbitrator for Large Crossbar Packet-Switches. *IEEE Transactions on Circuits and Systems* 36(7) (July 1989)
48. Timothy Brown, X., Liu, K.H.: Neural Network Design of a Banyan Network Controller. *IEEE Journal on Selected Areas in Communications* 8 (October 1990)
49. Badi, A., Akodadi, K., Mestari, M., Namir, A.: A Neural-Network to Solving the Output Contention in Packet Switching Networks. *Applied Mathematical Sciences* 3(29), 1407–1451 (2009)

# **Experimental Analysis of AODV, DSR and DSDV Protocols Based on Wireless Body Area Network**

Clement Ozugua Asogwa, Xiaoming Zhang, Degui Xiao, and Ahmed Hamed

College of Information Science and Engineering,

Hunan University, ChangSha 410082, China

acogugua@yahoo.com, zhangxm19712003@yahoo.com.cn,

jt\_dgxiao@hnu.cn, ahmedhmed@gmail.com

**Abstract.** Recently, more and more sensor network research is carried out in the area of remote and mobile health systems (mHealth) or Telemedicine applications. In these applications the sensors are centered in or on the proximity of the human body thereby forming a network called Wireless Body Area Network (WBAN) that is uniquely important for consideration. Hot problems currently researched in this network include reliability and energy efficiency of routing protocols. In this paper we present the analysis of three prominent routing protocols AODV, DSR and DSDV, based on the key characteristics of the WBAN using IEEE 802.15.4 MAC Protocol. In our experiments, packet delivery ratio and average end-end delay were used as the measure for the communication reliability and energy savings efficiency, while considering the mobility of the nodes, network communication range and the number of nodes. Based on the experimental analysis, both AODV and DSR protocols have good reliability, while DSDV performed very poorly with over 90 percent packet losses largely due to its high routing table overhead. Furthermore, in energy savings efficiency, DSR and AODV had similar performance while DSDV performed a little poorer.

**Keywords:** WBAN, IEEE 802.15.4, Sensor Network, AODV, DSR, DSDV.

## **1 Introduction**

WBAN is a type of sensor network that consists of intelligent, miniaturized, low-power sensor nodes attached on or implanted in the body which has the ability to establish wireless communication link one with the other. Usually the device consists of sensors and actuators [3]. The sensors have the ability to measure parameters such as heartbeat, body temperatures, gastrointestinal tract, neurological disorders and cancer detection etc and to record prolonged electrocardiogram (ECG) while actuators take specific actions according to the data they receive from the sensors or from other users. WBAN provides long term health monitoring of patients under natural physiological and psychological conditions and or without impacting on their natural activities. It can be an in-body or on-body network. WBAN nodes can

have different topological arrangements such as star, tree and mesh [9] which depending on the application can have nodes sometimes combined to process and transfer data to a central location or coordinator. The normal data is collected and processed by the coordinator and forwarded to systems such as Electronic Health Database, a Telemedicine Server, emergency health unit, and a medical ambulance etc. Unlike sensor networks, the relative movement of some parts of human body makes the classification of WBAN to take some characteristics of stationary sensor nodes and mobile sensors sometimes almost simultaneously. Also a body area network has the smallest coverage area. Usually this device [13] wrote are placed within 3m distance. Both wearable and body implant are limited in the number of nodes depending on application for humans and animals etc to ensure patients comfort, usually in the range 20-50 [3, 10]. The IEEE chartered a sub-working group within 802.15 (IEEE 802.15 Task Group 6 [TG6]) to bring a WBAN standard that provides medium access layer and physical layers which can be used for building applications [1]. The IEEE 802.15.4 is a low-power standard designed for low data rate applications [6], thus presently most WBAN applications are designed using IEEE 802.15.4 standard. WBAN uses the Wireless Medical Telemetry Services (WMTS), unlicensed Industrial, Scientific and Medical band, and Ultra-Wide band (UWB) and Medical Implant Communications Services bands for data transmission.

As stated earlier while sensor nodes measure set parameters from the human body example heart beat (ECG), etc; actuators take specific actions according to the data received from the sensors or through interaction with the user. Interaction with the user is usually through personal devices e.g. PDA or smart phones which act as sink for data from the wireless devices. Except for few on-demand traffics, most communication with the body sensors are simplex in nature, transmitting signal from the sensors to a sink. However [9] in his analysis showed that communication between implants should be through a central coordinator as peer to peer communication results in increasing path loss. In this paper we examined the performance of three prominent ad hoc routing protocols, AODV, DSR and DSDV over the IEEE 802.15.4 MAC protocol and examined its performance based on key criteria for WBAN such as reliability and delay. Section 1 introduced the general overview of WBAN; in section II we examined related works both in WBAN, the energy saving MAC protocol -IEEE 802.15.4 and related works involving the AODV(Ad hoc On-demand Distance Vector routing), DSR(Dynamic Source Routing) and DSDV(Destination-Sequence Distance Vector routing) protocols. In Section 3 we considered the metrics for our analysis and our simulation experiments in section 4. Finally in section 5, we rounded up with the conclusion and provide improvement areas this analysis can be extended as further research work.

## 2 Related Works

Body area networks are generally considered as special type of sensor network with focused requirements such as demand for reliability, energy efficiency and mobility support. The most common mechanism for reducing energy consumption is through

the MAC protocol design. The MAC protocol is used in controlling the power and duty cycling of the radio module and in reducing the average energy consumption of the sensor node by controlling the main sources of energy waste such as collision, idle listening, overhearing, and control packet overhead. The low-power mechanism plays an important role in the performance of a good MAC protocol for WBAN. The IEEE802.15.4 is a low-power standard designed for low data rate applications [6]. Some of the main reasons of selecting IEEE802.15.4 for WBAN are the low-power communication and support for low data rate, a characteristic for WBAN applications. [7] Investigated the performance of a non-beacon IEEE 802.15.4 for low upload and download rates and concluded that the non-beacon mode results in longer sensor life time for low data rate asymmetric WBAN traffic. Benoit Latre et al,[12] proposed a cross-layer communication protocol (CICADA) based on a tree structure. It uses a control subcycle and a data sub cycle to achieve low delay and energy efficiency; however it does not support traffic from the sink to the nodes. [14] Evaluated IEEE 802.15.4 over AODV with sink mobility and concluded that the sink node velocity should be less than 1m/s for obtaining acceptable performance under some predefined traffic loads and packet sizes using Random Trip Mobility Model. Campbell et al [19] demonstrated through simulations that IEEE 802.15.4 operates better with small data size at much shorter range while IEEE 802.11 performs better with high data rate, longer range and streaming bit rate.

To realize efficient communication in WBAN, techniques from Wireless Sensor Networks (WSNs) and ad hoc networks could be used however; due the specific characteristics of the WBAN most of the current protocols designed for sensor networks in 802.11 are not always suited in 802.15 [19]. Below is a list of some of the characteristic features of WBAN

1. Energy Efficiency: WBAN devices have limited energy resource with usually ir-rechargeable and unchangeable batteries especially for implants; low power consumption and losses are imperative.
2. WBAN devices are miniaturized without redundant parts in order to achieve the low form factor needed for body implants. Reliability of data and communication with low latency should be a must.
3. In order to cater for health issues, an extremely low power per node is needed to minimize interference as to cope with health concerns [11]. Each node is designed with low transmit power and highly prevented from interferences while co-existing with other WBAN.
4. The propagation medium of WBAN network is highly heterogeneous and very lossy as a result the waves are always attenuated considerably before they reach the receiver [2].
5. WBAN devices are used on human body which can sometimes be in high motion example a sprinter, slow motion during side work or at rest etc the topology can be highly varying.
6. The sensitive and personal nature of the data information carried makes it more demanding for reliability, low delay, highly secure with adequate confidentiality both external and internal.

7. Furthermore, these sensors perform different and independent functions, this means that there heterogeneity implies different levels of resource requirements from the network such as, data rates, power requirements, service priorities during emergency life critical messages etc.

In comparison with other sensor networks, WBAN is operated close to human body with communication range usually restricted to a few meters typical values being 2-3m [13].

Presently WBAN uses IEEE802.15.4 Mac protocol designed for low rate energy efficient Mac designs which support simple devices that consume minimal power and typically operate in the personal operating space (POS) of 10m or less [18]. A standard MAC protocol for WBAN is still being formulated by the subgroup IEEE 802.15.6 from the IEEE 802.15.4. Unlike traditional sensors, WBAN is similar to WSN but with few fundamental differences that requires close observation. Example the Node density in WBAN is usually fewer in most applications especially when used for health monitoring it is further limited in network area or communication range. Also, unlike WSN the node tasks in WBAN can be multiple with reliability depending on the nodes robustness whereas WSN has built in redundancies and performs dedicated tasks. The network topology of WBAN is highly variable due to body movement. The aggregate power demand is lower but with more difficult supply. The impact of data loss is very significant; as a result it requires measures to ensure Quality of service and real-time data delivery with low power technology as very essential. The protocol performances are statistically analyzed following the fundamental features of the WBAN. The result of the experiment will clarify each protocols based on the selected features primarily geared towards energy savings, ability to successfully deliver packets with minimal losses-Reliability, less delay, motion capable with few nodes and limited space with fast transmission time.

### **3 Metric for Analysis**

In order to experimentally examine the applicability of some prominent routing protocols for WBAN, the following metrics derived from the features listed above [for WBAN] are tested using Network Simulation 2 with WPAN extension and analysed. The protocols are AODV, DSR and DSDV. The first two are prominent on-demand routing protocols and share certain outstanding characteristics. Both discover routes only in the presence of data packets in need for a route to a destination while route information is stored in all intermediate nodes in the form route entries in AODV and route caches in DSR. DSDV however is a proactive protocol with frequent updates of routing tables regardless of need. The dynamic differences in these three prominent routing protocols will affect significantly the performances in all areas of application including energy saving MAC protocols thereby making further understanding of the dynamics of these protocols key to the analysis we are performing.

AODV structure has the following advantageous features. Routes are established on demand and destination sequence numbers are used to find latest route to destination thereby making connection set delay to be low. Because it does not make use of source

routing, it does not place additional overhead on data packets. It favours the least congested route instead of the shortest path and has support for unicast and multicast packet transmissions. However, AODV can lead to inconsistent routes if the source sequence number is very old and intermediate nodes have higher but not the latest destination sequence number (stale entries). It assumes that all nodes must cooperate without which no route can be established, therefore it is susceptible to various forms of attacks. Its network performance decreases as the network size grows.

DSR on the other hand has the ability to reduce network bandwidth overhead, conserve battery power and avoid the propagation of large routing updates because it does not use periodic routing messages. Routes are maintained only between nodes that need to communicate; with DSR route caching also helps to reduce route discovery overhead. It guarantees loop-free routing and rapid recovery when routes in the network changes. DSR is designed to be able to compute correct routes in the presence of asymmetric (uni-directional) links. The major draw back is that it is not scalable to large networks, usually efficient for mobile ad hoc networks with less than 200 nodes [21]. It also has more processing resources than most other resources.

DSDV is a table driven routing scheme based on Bellman-Ford Algorithm. It requires regular update of its routing tables which uses up battery power and small amount of bandwidth even when the network is idle. It reconfigures new sequence number for the network to converge whenever there is a topology change. Its packet delivery is low because it uses stale routes in cases of broken links. It is suitable for networks with small number of nodes [22]. To study the performance of these protocols for WBAN a number of qualitative and quantitative metrics can be used to study the behavior of these protocols in a simulated WBAN characteristic environment. The following are considered for this analysis.

- Node Mobility
- Network Communication Range and node density.

Other Protocol Performance characteristics include:

- Packet Delivery Fraction (PDF) as a measure of the routing protocols reliability.
- Average End-to-End Delay as a measure of the protocols' energy conservation efficiency.

The packet delivery fraction or ratio is defined as: The ratio of the number of packets received at the destination to the number of packets sent from the source.

$$PDF = \frac{\sum \text{Number\_of\_received\_data\_packets}}{\sum \text{Number\_of\_sent\_data\_packets}} \quad (1)$$

This parameter gives a measure of the reliability of the protocol under the specified condition. The greater the packet delivery ratio, the more reliable the network is.

Average Network Delay or End-End Delay is defined as: The average time delay for data packets from the source node to the destination:

$$\text{End-End Delay} = \frac{\sum (Time_{\text{packetsArrivedAtDestination}} - Time_{\text{packetsSentFromSource}})}{\text{TotalNumberOfConnectionPairs}} \quad (2)$$

Our analysis shall be drawn from a statistical result of the performances from the above based on the tested conditions of WBAN using NS-2 with WPAN extension [16].

## 4 Simulation and Analysis

We modified the NS-2 code with WPAN extension [15, 16] at the core and fine tuned several parameters where necessary to satisfy our simulation conditions. The performance is evaluated with respect to the following parameters in 802.15.4 and 802.11 Mac. We used the node-movement generator available in ns2 to run the setdest scenarios which defined our arguments in the simulation for setting the range, simulation time, speed range and topology boundary as well used the traffic-scenario generator to set the traffic limits.

The following graph defines the different simulation scenarios in both the 802.15.4 MAC and 802.11 MAC protocols.

From the graphs obtained, fig.1 shows that average packet delivery fraction is reasonably good in AODV and DSR when the network range and speed is 2m square and 0-5m/s respectively. DSDV has less than 3% success delivery of packets. Similarly retracing the experiment by keeping the number of nodes constant and varying speed and communication ranges, we have in fig.2 and fig.4 an average PDF of upwards of 95% in DSR and AODV, while DSDV remained consistently poor. DSR had better result in terms of reliability.

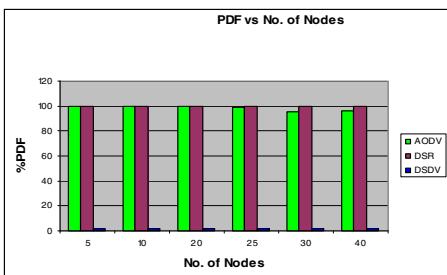


Fig. 1. 802.15.4 @ 2sqm range and 0-5m/s

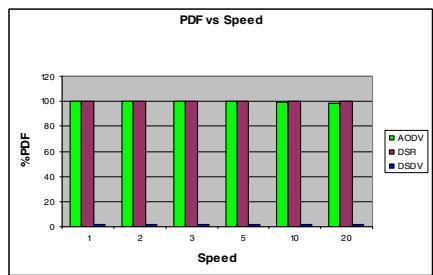


Fig. 2. 802.15.4@3m range, varying speed

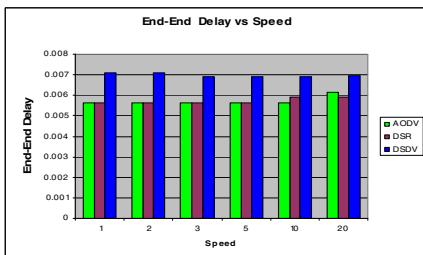


Fig. 3. 802.15.4@ 3m range and varying speed

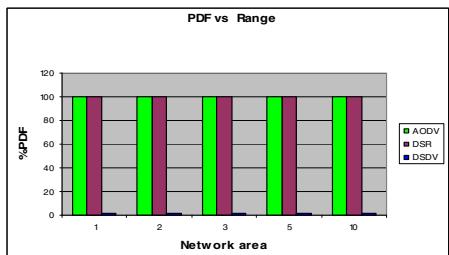


Fig. 4. 802.15.4 at varying range; max speed5m/s



Fig. 5. 802.15.4 @ constant node, 0-5m/s

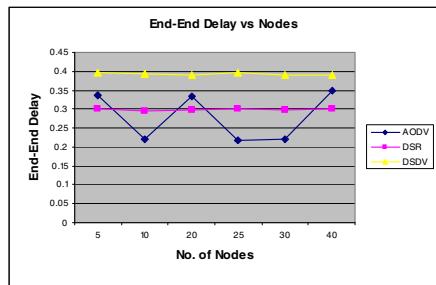
Fig. 6. 802.11 @ diff. nodes and 3m<sup>2</sup> range

Fig.3 and fig.5 shows the result of the End-End delay at various speeds and the communication ranges. The delay in AODV and DSR increased gradually as the speed increased to 20m/s (fig.3). Fig.5 shows DSR has higher energy requirement when the network range is 1m, the delay reduces as the communication range increases from 2-10m. The DSR performance at ranges less than 1m will be a nice experiment to further examine. The experiments were repeated using 802.11 MAC as shown in fig.6, fig.7 and fig.8. The result showed DSDV having good PDF and End-End delay better than DSR and AODVC in similar situation. The concluding result shows that both AODV and DSR have good packet delivery fraction (reliable in 802.15 MAC) at speed range 0-5m/s and 2-5sqm network communication range, with the number of nodes between 5 and 40 units.

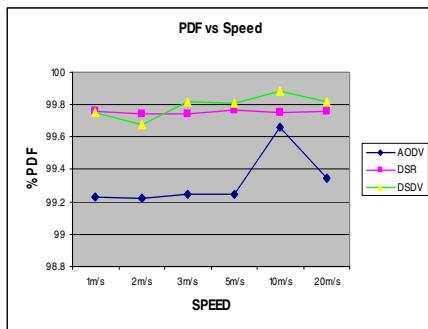


Fig. 7. MAC at varying range

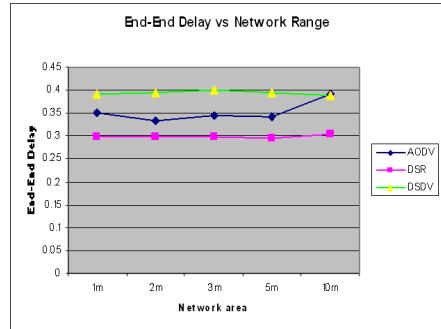


Fig. 8. Delay at varying Speed and constant node-in 802.11 MAC

While DSDV has over 90% packet delivery failure that can be attributed to high routing table overhead of the protocol given the limited packet size of 50 and CBR data flow rate of 0.1Mb as set in the 802.15 MAC. A repeat of the experiment using 802.11 MAC showed that DSDV had not the same high packet loss essentially showing that it is due to the protocol design (energy savings and low packet size of IEEE 802.15.4).

## 5 Conclusion

The performance of a routing protocol is usually measured with parameters such as Packet Delivery Fraction, End to End Delay etc in following the scenarios used in the experiment e.g. pause time, node speed, number of nodes and node communication range. In WBAN as in other energy saving protocols, these metrics are used to check the reliability, energy efficiency and resourcefulness, including ease of operation. In this paper we carried out analysis of three prominent routing protocols; one proactive and two reactive protocols. The results have been reflected in graphs. It has been analyzed that both AODV and DSR are good in performance in the tested cases. While DSR is steadier and very effective in performance at very small number of nodes, our experiment was limited to 40 nodes following current design for acceptable number of nodes in WBAN. However, according to previous research works, the performance of DSR changes as the number of node increases and fails when the number reaches upwards of 100 [21].

While improvements have been made in these protocols our analysis did not consider improvements in them and will form part of our future work. As future developments in WBAN are expected both in the application and design, several nodes may be required, our future work will include analysis involving higher number of nodes.

## References

1. Bilstrup, K.: Preliminary Study of Wireless Body Area Networks. Technical Report IDE0854 (August 2008)
2. Amjad, K., Stocker, A.J.: Impact of slow and fast channel fading and mobility on the performance of AODV in ad-hoc networks. Dept. of Eng., Univ. of Leicester, Leicester, UK (November 2010)
3. Latre, B., Brean, B., Moerman, I., Blondia, C.: A survey on Wireless Body Area Networks. *Journal of Wireless Networks* 17(1), 1–12 (2011)
4. Ullah, S., Higgins, H., Braem, B., et al.: A Comprehensive Survey of Wireless Body Area Networks on PHY, MAC and Network Layers Solutions. Springer Science + Business Media, LLC (2010)
5. Sayranfin-Pour, K., Yang, W.-B., Hagedorn, J., Terrill, J., Yazdandoost, K.Y.: A statistical path Loss model for medical implant communication channels. In: The Proc. of 2009 IEEE 20th International Symposium on Indoor and Mobile Radio Communications, pp. 2995–2999 (2009)
6. IEEE std 802.15.4.: Wireless medium access control (MAC) and physical layer (PHY) specifications for low data rate wireless personal area networks (WPAN). IEEE, Piscataway (2006)
7. Timmons, N.F., Scanlon, W.G.: Analysis of the performance of IEEE802.15.4 for medical sensor body area networking. In: The Proc. of the First Annual IEEE Communication Society Conference on Sensor and Ad Hoc Communications and Networks (IEEE SECON 2004), pp. 16–24 (2004)

8. Tang, Q., Tummala, N., Gupta, S.K.S., Schweiber, L.: Communication scheduling to Minimize thermal effects of implanted biosensor networks in homogenous tissue. *IEEE Transactions on Biomedical Engineering* 52(7), 1285–1294 (2005)
9. Ullah, S., et al.: A Comprehensive Survey of Wireless Body Area Networks on PHY, MAC, and Network Layers Solutions. *J. Med. Syst.* (July 25, 2010), doi:10.1007/s10916-010-9571-3
10. Otto, C., Milenkovic, A., Sanders, C., Jovanov, E.: System architecture of a wireless body area sensor network for ubiquitous health monitoring. *Journal of Mobile Multimedia* 1(4), 307–326 (2006)
11. Rui, P.J., Foster, K.R.: Heating of tissue by near-field exposure to a dipole: a model analysis. *IEEE Trans. Biomed. Eng.* 46(8), 911–917 (1999)
12. Latre, B., et al.: A Low-delay Protocol for Multihop Wireless Body Area Networks. Ghent University–IBBT-IMEC-Dept. of Information Technology-IBCN
13. Kwak, K.S., Ameen, M.A., Kwak, D., Lee, C., Lee, H.: A study of the Proposed IEEE802.15 WBAN MAC Protocols. In: *ISCIT 2009*, Inha University, Telecommunication Research Institute (ETRI), Korea (2009)
14. Gowrishankar, S., Basavaraju, T.G., SubirKumarSarkar: Simulation Based Analysis of Mobile Sink Speed in Wireless Sensor Networks. In: *Proceedings of the World Congress on Engineering and Computer Science 2010*, vol. 1 (2010)
15. Parc, X., Fall, K., Varadhan, K.: The VINT Project, U.C. Berkeley, LBL, USC/ISI, The ns Manual (formerly ns Notes and Documentation)<sup>1</sup> (January 2009)
16. Information Sciences Institute, The Network Simulator NS-2,  
<http://www.isi.edu/nsnam/ns/>
17. Zheng, J., Lee, M.J.: A Comprehensive Performance Study of IEEE 802.15.4. In: *Sensor Network Operations*, ch. 4, pp. 218–237. IEEE press, Wiley Inter Science (2006)
18. Zheng, J., Lee, M.J.: Will IEEE 802.15.4 Make Ubiquitous Networking a Reality? A Discussion on a Potential Low Power, Low Bit Rate Standard. *IEEE Communication Magazine*, 140–146 (June 2004)
19. Campbell, C.E.-A., Loo, K.-K., Kurdi, H.A., Khan, S.: Comparison of IEEE 802.11 and IEEE 802.15.4 for Future Green Multichannel Multi-radio Wireless Sensor Networks. *International Journal of Communication Networks and Information Security (IJCNIS)* 3(1), 96–103 (2011)
20. Jonanov, E., et al.: A wireless body area network of intelligent motion sensors for computer assisted physical rehabilitation. *Journal of NeuroEngineering and Rehabilitation* (March 2005)
21. Taneja, S., Kush, A., Makkar, A.: Experimental Analysis of DSR,AODV using Speed and Pause time. *International Journal of Innovation, Management and Technology* 1(5), 453–458 (2010)
22. Perkins, C.E., Bhagwat, P.: Highly dynamic destination-sequenced distance vector routing (DSDV) for mobile computers. In: *Proc. ACM SIGCOMM 1994*, London, UK, pp. 234–244 (October 1994)

# OCTBR: Optimized Clustering Tree Based Routing Protocol for Wireless Sensor Networks

Jian Zhang, Yu Xie, Dandan Liu, and Zhen Zhang

Computer School, Wuhan University, Wuhan, Hubei, China

{jzhang, csliudd}@whu.edu.cn, {whucsxy, zzhang}@163.com

**Abstract.** Routing protocol plays the vital role for information acquisition in wireless sensor networks. This paper proposes a novel routing protocol, called Optimized Clustering Tree Based Routing (OCTBR) protocol. In this protocol, the nodes capability and network environment are taken as the major considerations for the protocol design. The ratio of the cluster head is set according to the scale and density of the network, and the clustering mechanism is implemented based on some factors such as node's energy and the distance between nodes. When a large number of nodes in WSN become invalid, the network starts fast local routing repair algorithm to restore normal routing as soon as possible. The results of the simulation show that, compared with the existing tree based routing protocols, the OCTBR protocol can significantly increase the network throughput, packets reception rate, and finally reach the purpose to prolong the network lifetime.

**Keywords:** Tree based routing protocol, clustering, routing recovery.

## 1 Introduction

Wireless sensor network is generally used to collect sensing data for the internet of things. WSN usually includes sensor nodes and sink nodes, which are distributed in the sensing area randomly. Routing protocols in wireless sensor networks are the main function of the network layer. To design a routing protocol with higher communication efficiency and quality, lower energy consumption, and longer lifetime is always one of the core problems in the WSN. There are many challenges in the design and implementation of an efficient routing protocol. Firstly, the system is limited by the energy capacity of sensors nodes. Sensors are powered by batteries and it may not be possible to recharge or replace them in the real situation. Secondly, WSN is fragile due to the vulnerability of sensor nodes. WSN is often placed in some dangerous areas, such as battlefield or disaster area. Sensor nodes usually cannot resist physical environmental influences. If some nodes in the dangerous area are damaged, the data of the area covered by these damaged nodes will not be collected correctly. Moreover, it may prevent other normal nodes from sending message to base station because those damaged nodes are elected as route intermediary. Thus, we should consider the situation that single or multiple nodes fail at the same time.

The traditional routing protocols that based upon flooding or clustering method have drawbacks. Flooding-based strategy usually brings huge energy consumption,

even the information explosion and network congestion. There are two typical clustering-based protocols, LEACH and PEGASIS. LEACH algorithm improves the performance over the plane algorithms, but the election of cluster head still has many defects. PEGASIS protocol makes improvement on the link of LEACH. The election of cluster head is dynamic, and each sensor node chooses the nearest neighbor node as the next hop according to their mutual position relationship. But the problem is that it may cause longer chain and longer delay of data transmission.

Considering the energy efficiency and other affecting factors of routing, we propose an Optimized Clustering Tree Based Routing (OCTBR) protocol. The election of cluster head is based on the available energy of nodes and the relative positions between nodes. After the election completes, a cluster head tree is created and the distance from each cluster head to the base station is minimum. Meanwhile, considering the situation when a number of nodes fail, we adopt a recovery strategy to rebuild transmitting route in the failure zone by speeding up the routing information propagation. The simulation results show that the OCTBR protocol can significantly increase the energy efficiency, prolong the network lifetime and achieve good load balance over the LEACH and PEGASIS.

The rest of this paper is organized as follows. Section 2 gives the related work. Section 3 introduces the process of cluster construction by OCTBR protocol, including cluster head election, cluster head tree construction and recovery strategy for routing optimization. Section 4 gives the simulation of the proposed protocol, and compares it with the other two existing protocols LEACH and PEGASIS. Finally we conclude this article in section 5.

## 2 Related Work

In order to adapt to different needs of the WSN applications, researchers have proposed various routing protocols. The typical one is the flooding-based protocol. But they cannot overcome the problems of message imploding. Another kind is the location-based protocol, such as GEAR[1]. It takes greedy algorithm which is a local optimization algorithm, but the situation of “void area” will decrease the success rate. In another similar protocol MECN[2], node sends data to its forwarding area. The drawback is that MECN assumed that every pair of nodes can communicate with each other directly. However, this assumption is not realistic at all. It is suitable for the networks where nodes are fixed or less mobile.

LEACH and PEGASIS are representatives of hierarchical routing protocols in clustering method. LEACH [3] adopts data fusion, which can reduce the network traffic overhead effectively and achieve the node energy load balance. But it has some problems when applied in real application: (1) LEACH is not suitable for large-scale WSN as it assumes that each sensor node in the network can communicate with the sink directly. (2) It sets the number of the cluster heads to be 5% of the total nodes. But it cannot ensure the distribution of the heads. (3) The network lifetime will be greatly reduced if the selected heads are out of energy. The PEGASIS [4] makes some improvement over the LEACH. However, it still has the following shortcomings: (1) PEGASIS takes much system overhead to update the location and energy information

of all other nodes. (2) It constructs one chain for the cluster. The only head of the chain plays the key role for the routing. Any failure happened on the head will cause the failure of the entire network. (3) It will lead to excessive delay of data transmission for the nodes far away from the chain head.

Wang Dong et al.[5]proposed a fast self-recovery method based on the distance vector routing protocol, called FS-DVP. The main idea is that each node calculates the possible next hop for each destination in advance and stores them. It can effectively improve the stability and efficiency of the network. The LEACH-C, LEACH-F protocol proposed by Heinzelman et al.[6] are based on LEACH. Based on node's position, remaining energy, and the distance, the base station selects the appropriate clusters and their heads. The DCHS protocol proposed by Handy [7] avoids choosing a fixed node as the cluster head. However, the energy consideration makes the algorithm more robust.

### **3 The Optimized Clustering Tree Based Routing (OCTBR) Protocol**

In order to avoid the problems stated in section 2, we propose an Optimized Clustering Tree Based Routing (OCTBR) protocol. OCTBR adopts a fixed clustering method. Nodes will not change clusters during the entire life time once they select the appropriate cluster to join in. When the elected cluster head is too low in energy or destroyed, a new cluster head election will be taken within the same cluster. Instead of setting a fixed rate for cluster head, OCTBR calculates the optimal number according to nodes' remaining energy. Later, all cluster heads will form a cluster head tree. All messages are transmitted from sensor nodes to each cluster head, and finally transmitted to the root, i.e., the base station, following the cluster head tree structure.

#### **3.1 The Cluster Head Election**

OCTBR protocol adopts the following assumptions. The density of nodes is similar to $\lambda$ Poisson distribution. Sensors and base station are static, and the base station has unlimited energy and is far from sensors. Cluster head receives data from the member nodes in TDMA mode, integrates all data and transmits them to the base station.

$n$  represents the total number of nodes in wireless sensor network. In the initialization stage, each node in the network transmits the message to its neighbors, including its own state, remaining energy and location coordinates information. Each node receiving the data from its neighbors will sends feedback. At the same time, it counts the number of its neighbors, the remaining energy  $E$  of each neighbor, and the distance  $D$  to each neighbor. According to a function  $T=F(D,E)$  with  $E$  and  $D$  as key parameters, each node decides whether to be a cluster head. After that, each cluster head sends a signal to their neighbors, and each node joins the nearest cluster. The cluster election details are demonstrated in the following sections.

Based on the energy consumption model used by Heinzelman et al[8], which is a famous model used for wireless sensor networks, and following a numerous calculation, we find that the best probability  $p$  of cluster head can be calculated as:

$$p = \sqrt{\frac{10\lambda}{0.0061632025n^2}} \quad (3.1)$$

where  $\lambda$  is the density parameter of Possion distribution, and  $n$  is the number of nodes in the network. The prerequisite for a sensor node  $i$  to be a cluster head is that its value  $T(i)$  is within the top  $p*n$ . That is

$$T(i) = F(D, E) = \frac{f_E}{f_D} \quad (3.2)$$

where  $f_D$  represents the distance function, and  $f_E$  represents the energy function. According to the formulation above, the  $T(i)$  of a node is inversely proportional to the relative distance between nodes, and proportional to its own remaining energy.

To determine the distance function  $f_D$ , we use a parameter  $C_i$ , which is the number of neighbors that are within the transmission range of the  $i$ -th node. From the formulation 3.1, we can determine the optimal percentage of the number of cluster heads is  $p$ . In other words, it is best to recruit about  $1/p$  nodes in each cluster. If the members in a cluster are too many, the cluster head will not work well; but if the number is too small, it will increase the energy consumption for cluster heads. As a result, we define the distance function as follows:

$$f_D = \alpha * (1 - (\frac{ci - 1/p}{1/p})^2) \quad (3.3)$$

To determine the energy function  $f_E$ , we set  $E_{cur}$  to denote current energy of a node. So we have:

$$f_E = (\beta * E_{cur})^2 \quad (3.4)$$

To determine the values of  $\alpha$ ,  $\beta$ , we can use the formulations 3.5 and 3.6 below. In the initial stage of network, the initial energy of each node is the same. Position can be the only factor to determine whether to be a cluster head or not. For example, the possibility for a node in the geometric center of the cluster region to be a cluster head is greater than the nodes in other positions. Thus, location parameter  $\alpha$  and the number of survival nodes in cluster have direct relationships with relative distance. That is

$$\alpha = \frac{\sum_{i=0}^m d(c, i)}{m} \quad (3.5)$$

Where  $m$  is the number of survival nodes, and  $d(c, i)$  is the relative distance between current node  $c$  and each survival node  $i$ .

After the wireless sensor network runs for a period of time, the distance between nodes has less influence on the election of cluster heads, and each node's remaining energy plays a leading role. Thus energy parameter  $\beta$  has a greater impact to the network lifetime. We assume that network lifetime is  $T$ , the current time is  $t$ , then we have

$$\beta = \frac{ti}{T} \quad (3.6)$$

Put  $\alpha$ ,  $\beta$  into formulation 3.3 and 3.4 respectively, we can get  $f_D$  and  $f_E$ . Put  $f_D$  and  $f_E$  into formulation 3.2, we can get the  $T(i)$  of each node  $i$ .

According to each node's  $T(i)$ , OCTBR can decide the clusters and their heads. Firstly, a threshold  $T$  is predefined according to the number and distribution of nodes. When the  $T(i)$  of a node  $i$  exceeds this threshold (i.e.,  $T(i) > T$ ), the node  $i$  is considered to be the cluster head candidate. Note that the energy consumption is proportional to the distance between two nodes. Thus in this primary election of cluster heads, OCTBR will choose the node that is nearest to the geometric center of each cluster region to be a head as much as possible. After choosing the entire cluster head candidates, OCTBR determines the number of heads by the optimal cluster head ratio  $p$ . The nodes whose  $T(i)$  is within the top  $p*n$  number in the ordering of  $T(i)(0 < i < n)$  are finally selected as the cluster heads.

Later, each cluster head makes an announcement. All nodes receiving this announcement will calculate its distance relative to the announcement sender, and choose the cluster having the nearest cluster head to join in. At this moment, the clusters construction is finished. Each cluster member will wait for the schedule information from the cluster head for future transmission.

### 3.2 Cluster Head Tree Construction

After the completion of the cluster building, the base station (i.e., sink) broadcasts messages to the surrounding cluster heads by power  $P$ . Note that when a node is elected to be the cluster head, it will use power  $P$  to communicate with other heads. Set the base station to be the root, the cluster heads which can receive messages from the base station are its child nodes. After receiving messages, the heads will send feedback to base station and join the cluster head tree. These nodes which become the members of cluster head tree broadcast messages to the surrounding cluster heads similarly. The process repeats until all the cluster heads are added to the cluster head tree. During the broadcasting process, the member that does not receive messages from any other cluster heads within time  $t$ , becomes the leaf node of the tree and will stop sending any messages.

A broadcasting message always includes the distance between the sending and receiving nodes. According to the message, each node can compute the shortest route from the base station to itself. As a result, if a cluster head receives messages from multiple parent heads, it will choose the node that can comprise the shortest route as its parent node. This can ensure that the path from each cluster head to the base station is the shortest, and minimize the energy consumption during the transmission. After the cluster head tree is built, the entire network topology structure is completed.

### 3.3 Fast Routing Recovery

As the scale of wireless sensor network expands, the maintenance for nodes becomes more and more difficult. The failures of nodes will affect the function of network. In

order to save energy, existing protocols usually set the interval between packets transmission to be quite long. Therefore, if some nodes fail, it takes a long time to convince the failure before taking strategy to recover. To make an improvement, we proposed a fast routing recovery method for OCTBR optimization. It shortens the transmission interval when necessary, speeds up the packet dissemination, so that it can fast recover from the failure. The value of the interval can be set according to some pre-experience. The fast routing recovery will be triggered when the failure node is a cluster head node. Specifically, a cluster head node whose energy or packets' reception rate is below a threshold will be defined as a failure node. Then a new elected cluster head will negotiate with the original child nodes, and starts to receive data from them later on.

## 4 Performance Evaluation

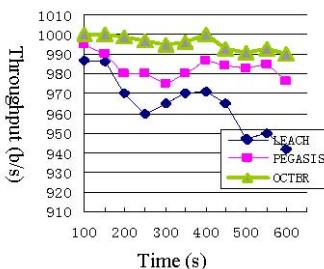
In this section, we will study the performance of the proposed OCTBR protocol through simulations. We use ns-2.27 for the simulation.  $n$  sensor nodes whose initial energy is 2J are distributed in a 100\*100 square plane. Each node's position is fixed. Set the amount of energy dealing with per bit data by transmitter or receiver to be Eelec=12nJ/bit. The energy consumption is 2pJ / bit when signal amplifying circuit transmits per bit data to per unit area. The energy consumption is 0.0005pJ/bit when the sensor node transmits per bit. When cluster heads are processing signal and fusing data, the energy consumption per bit is EDR=1nJ/bit/signal. The energy consumption of node in sleep state is 0.0001pJ/S. The size of each data packet is fixed to be 512bit. Links use tcp to connect. The bandwidth is 1MB/s. 250 sensor nodes send data to base station every second respectively. The simulation time T=600s. The fast routing recovery starts at 200s. The method is to set the invalid node's energy to 0. Each of the following results is an average of 10 independent experimental results. The performances of OCTBR, LEACH and PEGASIS are compared in the following terms: the network's throughput, packet reception rate and lifetime.

### 4.1 Throughput of Network

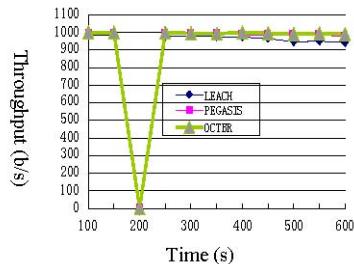
We first show the network throughput of OCTBR, LEACH and PEGASIS respectively. The throughput of the overall network with or without routing failure are shown in Figure 1-2. Thanks to the fast recovery strategy, OCTBR recovers to the normal state faster than the other two protocols when some nodes become invalid.

### 4.2 Packet Loss Rate of Network

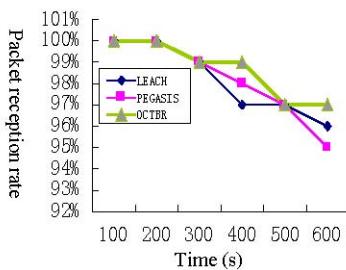
Figure 3 shows the average packet reception rate about OCTBR, LEACH and PEGASIS respectively during period T. Figure 4 shows the average packet reception rate when routing failure occurs. When routing failure occurs, because invalid cluster head is close to the sink, the child nodes of invalid node cannot transmit data to the sink, which leads to the packet reception rate is almost 0. Figure 4 shows, the OCTBR protocol can restore routing in a short time and effectively increase the packet reception rate.



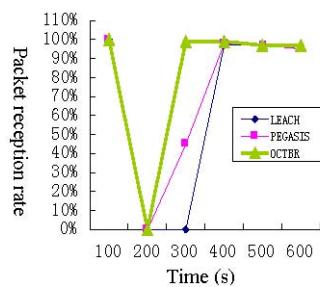
**Fig. 1.** The throughput of network without node's failure



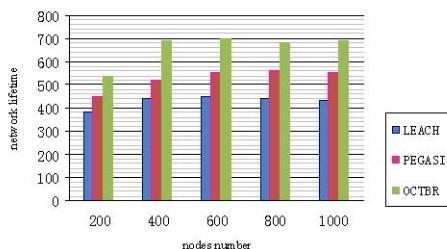
**Fig. 2.** The throughput of network with node's failure



**Fig. 3.** The packet reception rate without routing failure



**Fig. 4.** The packet reception rate with routing failure



**Fig. 5.** The lifetime of network changes with the number of nodes in the three protocols

### 4.3 The Lifetime of Network

We change the network distribution by increasing the number of the nodes from 200 to 1000 with increment in 200. The lifetime of network is shown in Figure 5. As shown in the figure, we can see that the lifetime of network by OCTBR protocol is longer than that by the other two protocols. The main reason is because in the initial stage, fast routing recovery algorithm constructs the network topology in a short time. Later, it doesn't have to search new cluster head node each round. As a result, a large amount of energy can be saved which lead to a much longer lifetime.

## 5 Conclusion

This article proposes a novel protocol named OCTBR that can achieve high energy efficiency in routing. OCTBR makes improvements on the calculation of the optimal number of cluster heads, the election of cluster heads and the construction of the cluster heads tree. In addition, OCTBR applies fast routing recovery algorithm, which can quickly recover from routing failure by accelerating information transmission frequency. Compared with the traditional routing recovery methods, OCTBR can not only reduces the routing recovery time, but also improves the utilization rate of energy and prolongs the lifetime of the network significantly.

**Acknowledgments.** This work was supported in part by the Fundamental Research Funds for the Central Universities (No. 3101049), the Natural Science Foundation of HuBei Province of China (No. 2011CDB458), and the Research Fund for the Doctoral Program of Higher Education of China (No. 20110141120038).

## References

1. Yu, Y., Govindan, R., Estrin, D.: Geographical and Energy Aware Routing: A Recursive Data Dissemination Protocol for Wireless Sensor Networks. UCLA Computer Science Department. Technical Report UCLA/CSD-TR-01-0023 (May 2001)
2. Xu, Y., Heidemann, J., Estrin, D.: Geography-informed energy conservation for ad hoc routing. In: 7th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom 2001), Rome, Italy (July 2001)
3. Heinzelman, W., Chandrakasan, A., Balakrishnan, H.: Energy-efficient communication protocol for wireless sensor networks. In: Hawaii International Conference System Sciences, Hawaii (January 2000)
4. Lindsey, S., Raghavendra, C.: PGEASIS: Power-Efficient Gathering in Sensor Information Systems. Computer Systems Research Department, The Aerospace Corporation, P.O. Box 92957, Los Angeles, CA (September 2001)
5. 王东,李娜,王文艳,吕格莉:节点密度对优化 Ad Hoc网络生存期的影响的研究 19(4), 234–240, 云南民族大学学报 (2010)
6. Heinzelman, W., Chandrakasan, A., Balakrishnan, H.: An application-specific protocol architecture for wireless microsensor networks. IEEE Transactions on Wireless Communications 1(4), 660–670 (2002)
7. Handy, M.J., Haase, M., Timmermann, D.: Low energy adaptive clustering hierarchy with deterministic cluster-head selection. In: 4th IEEE Conference on Mobile and Wireless Communication Networks, pp. 368–372 (2002)
8. Heinzelman, W.R., Chandrakasan, A., Balakrishnan, H.: An Application-Specific Protocol Architecture for Wireless Microsensor Networks. IEEE Transaction on Wireless Connections 1(4), 60–67 (2002)

# An Intelligent Irrigation System for Greenhouse Jonquil Based on ZigBee Wireless Sensor Networks

Zongyu Xu<sup>1</sup>, Baodong Lou<sup>2,\*</sup>, Guangcheng Shao<sup>3</sup>

<sup>1</sup>College of Computer and Information, Hohai University, Nanjing, 210098, China

<sup>2</sup>Engineering Training Center, Hohai University, Nanjing, 210098, China

<sup>3</sup>Institute of Water Conservancy and Hydroelectric Power, Hohai University,  
Nanjing, 210098, China

xbg7729@sina.com, loubd@126.com, 470203931@qq.com

**Abstract.** This paper analyzes the present situation of agriculture development in our country as well as the unique advantages of ZigBee wireless sensor network. By analyzing the intelligent irrigation system overall structure, the hardware design, with ZigBee wireless sensor network transceiver CC2530 being the core component has been clarified. On the basis of hardware, the software structure has been designed. Analysis of the growth characteristics and water requirement of Jonquil has been made. In the facility agriculture and environment experimental field of Hohai University Water Park, an experiment has been carried on the greenhouse jonquil irrigation, as an example of the design of irrigation system. The communication effect is satisfying; the desired effect has been basically achieved. Applying the intelligent irrigation system based on ZigBee Wireless sensor networks designed in this paper to greenhouse planting is of great significance to improve the condition of agricultural water utilization.

**Keywords:** ZigBee, intelligent irrigation, greenhouse jonquil cultivation.

## 1 Introduction

Chinese agriculture is unique with geographic dispersion, crop diversity, species diversity, complex factors. It is one of the areas that influenced by the environment most obviously. Therefore, timely collection and rational utilization of the agricultural environmental information data is significantly important. Being a traditional big agricultural country, China should learn from agriculture developed countries, to improve the utilization of irrigation water. Actively promoting agricultural modernization is our country's national condition requirement. Modern agriculture can not development without the intelligent irrigation system. With the progress of science and technology, the rapid development of wireless sensor network provides a new development opportunity for intelligent irrigation.

ZigBee, as a new international standard of short distance wireless communication protocol, is a kind of technical proposal between Wireless Markup technology and

---

\* Corresponding author.

Bluetooth technology [1]. Compared with other wireless communication protocol, it has features of a short distance, low rate, low complexity, low power consumption, low data rate and low cost. It is mainly used in the near distance wireless transmission.

The ZigBee protocol stack system structure is based on the standard seven open system interconnect (OSI) model, the IEEE802.15.4-2003 standard defines the following two layers: physical layer and medium access layer. Network layer, application convergence layer and application layer are formulated by the ZigBee alliance [2].

ZigBee wireless sensor network is the combination of sensor technology, communication technology and computer technology. It can collecting, processing and transmitting information [3]. With the development of sensor technology and the popularization of network technology, sensor network is used more and more widely. ZigBee technology can format network of star, tree and network structure automatically in equipments. It is widely used in national security, environmental monitoring, traffic management, modern logistics, warehousing management, medical care, intelligent home furnishing and other fields, and has good market prospect. Combined with the advantages of wireless communication, sensor network will have a broader development prospect and broader range of applications. To achieve a omnipresent sensor network is of great research significance and application value.

ZigBee technology makes up the vacancy for the low power consumption, low rate, short distance wireless technology standard [4]. With a broad space for development and broad application prospect, it is especially suitable for the construction of omnipresent sensor networks. Taking the greenhouse jonquil as an example, this paper discusses the application of ZigBee wireless sensor technology in precision agriculture water-saving irrigation.

## 2 The Overall Framework of the Intelligent Irrigation System

The design of the overall framework of intelligent irrigation system is made first. The overall system is shown in Fig.1. In terms of hardware, intelligent irrigation system is composed of a controller, a router, data acquisition nodes, the control nodes and irrigation device. The system operates regularly in Fig.2. In terms of software, the ZigBee sensor network programming is completed in the IAR for MCS-51 integrated development environment [5].

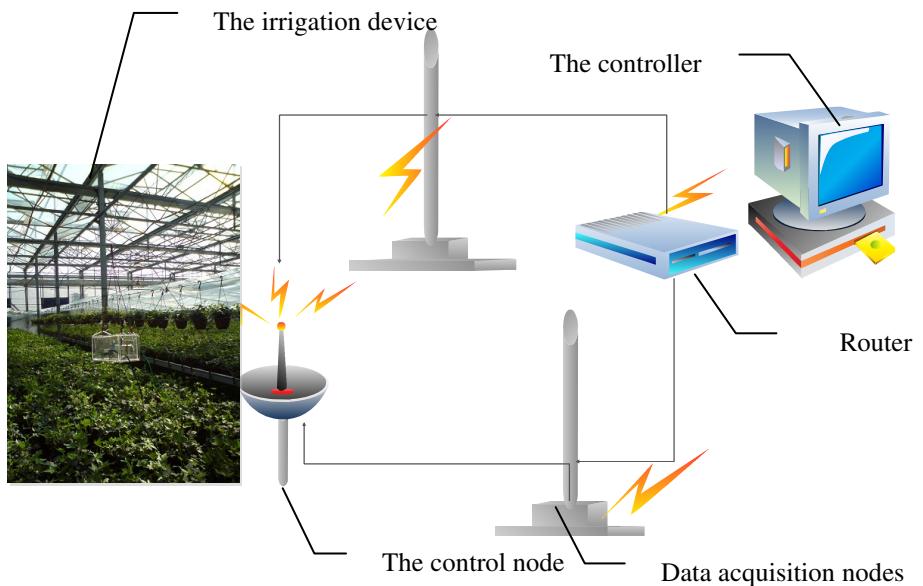
The controller, using ARM processor, is used to improve and expand the whole intelligent irrigation system performance.

Router is used to coordinate the intelligence operation of the whole system, in order to ensure the stability of the system. It is composed of ZigBee wireless sensor network transceiver module, power supply, external expansion memory.

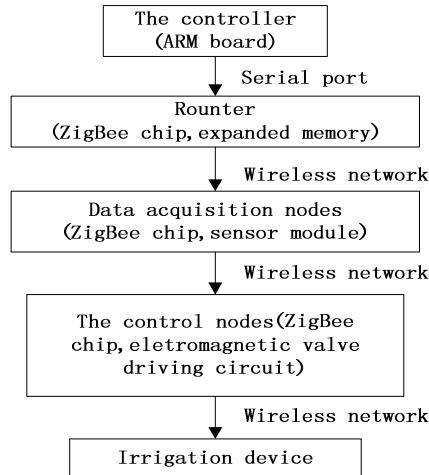
Data acquisition nodes, being the key part of the system, in addition to the ZigBee wireless sensor network transceiver, also set up various types of sensor module for the soil humidity, light, air temperature and humidity information collection.

The control node, through the electromagnetic valve drive circuit, the collected information is applied to the control of irrigation device timely.

The irrigation device is used for crop irrigation.



**Fig. 1.** The overall framework of the intelligent irrigation system



**Fig. 2.** The system frame diagram

### 3 Hardware Design

In this system, ZigBee wireless sensor network transceiver is the core component, its composition is particularly important. After careful selection, we choose CC2530 ZigBee chip produced by the TI company. CC2530 is used for 2.4-GHz IEEE 802.15.4, ZigBee and RF4CE application in a real system on chip (SoC) solution. It can build a strong network node with very low total cost of material. CC2530 combines the leading RF transceiver with excellent performance, the industry standard enhanced 8051 CPU in system programmable flash memory, 8-KB RAM and many other powerful features. CC2530 has different operation modes, making it especially suitable for ultra low power requirements of the system. Operation modes change in a short time to further ensure the low energy consumption. Wireless sensor network node hardware structure is shown in Fig.3.

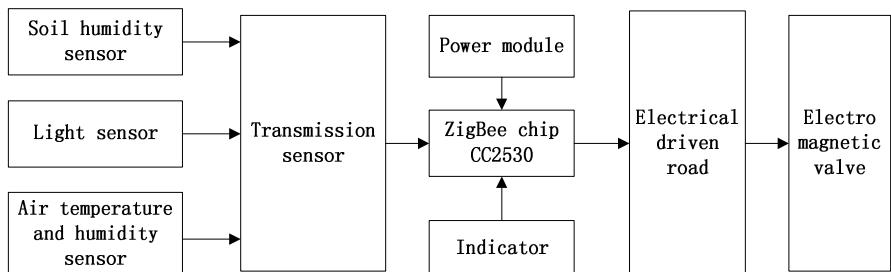


Fig. 3. Wireless sensor network node hardware structure diagram

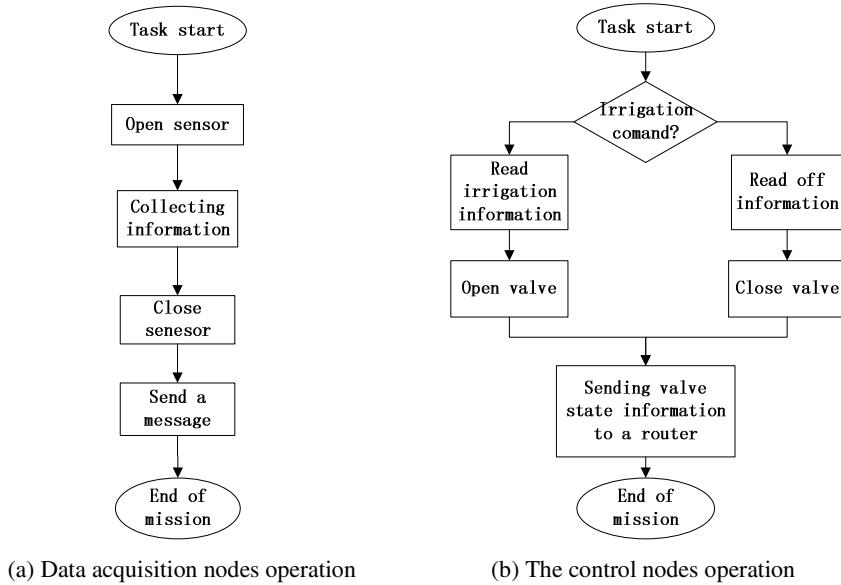
### 4 Software Design

#### 4.1 Software Development Environment

The system's development environment is IAR for MCS-51 that is provided by TI and is also convenient for programming CC2530[6]. IAR for MCS-51 is convenient to operate and connect and simple to learn. It is also the best and most practical tools to study Zigbee products. Through the USB interface connected to the computer, it has the code download, online debugging, breakpoint, single-step, observed variables, register observation functions, so as to realize the real-time online simulation and debugging of CC2530 series wireless single chip microcomputer. The development kit templates can help beginners and design staff finish the rapid assessment and a variety of Zigbee application development.

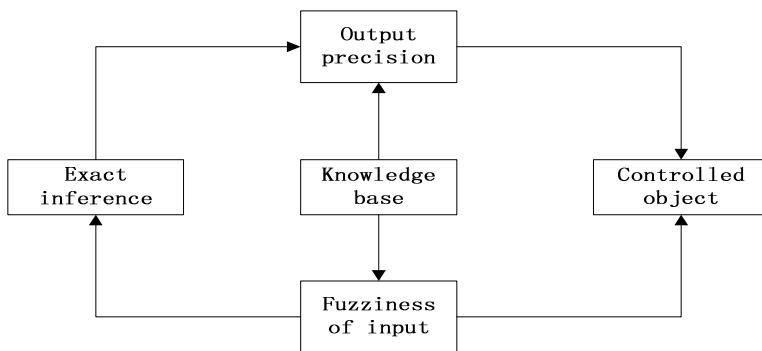
#### 4.2 Node Operation Process

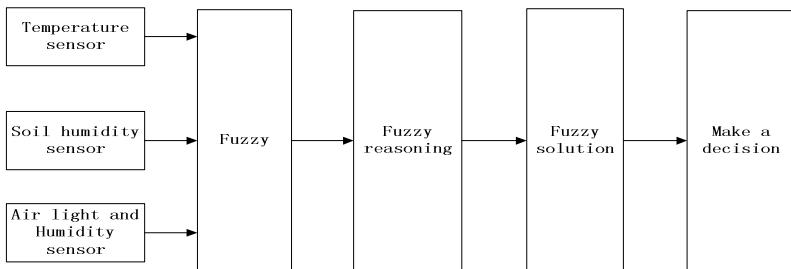
Data acquisition nodes operates as the order in Fig. 4a. The control nodes operates as the order in Fig. 4b.

**Fig. 4.** The operation process of nodes

### 4.3 Control Process

In the intelligent irrigation system, we apply the thought of fuzzy control in Fig.7. Fuzzy control theory is first proposed by the famous American scholar of University of California professor Zadeh• L• A in 1965, which is based on fuzzy mathematics, using linguistic rules representation methods and advanced computer technology. It is an advanced control strategy made by fuzzy reasoning decision [7, 8]. The basic structure of fuzzy is shown in Fig.5 and the control process in shown in Fig.6.

**Fig. 5.** Basic structure of fuzzy

**Fig. 6.** Control flow diagram

## 5 Greenhouse Wireless Sensor Network

In the facility agriculture and environment experiment field of Hohai University water park, we has made experiments on greenhouse.

Jonquil is succulent plant and favors warmer temperature. Its suitable growth temperature is lower than 30 degrees in the daytime and higher than 18 degrees at night. Too high or too low temperature will both affect the normal growth.

Jonquil propagates mainly by cottage. The temperature between 20 and 25 degrees suited to its roots growth most. Jonquil is fit well drained soils. Artificial cultivation soil with peat, vermiculite, perlite is given priority. The culture soil pH value preferably is between 5.8 and 6.2. Jonquil's water demand is low, it should be watered before noon and be kept dry before the night. The growth situation of jonquil is shown in table 1.

**Table 1.** The growth situation of jonquil

	Seedling stage	Growing stage	Mature stage
Required development time	About 2 months	About 3~4 weeks	About 4 months
The optimum soil moisture (%)	45~55	55~65	40~50

Note: The optimum soil moisture refers to the volume of water on soil.

Through the application of ZigBee wireless sensor network in greenhouse jonquil irrigation, we can observe the effects of intelligent irrigation on jonquil's quantity and quality.

In our experiment, the soil moisture sensor is the Hangzhou HuiEr corporation's based on time domain reflection method for measuring FDS-100 soil humidity sensor, the light sensor is TAOS company's TSL2550 digital light sensor, the air temperature and humidity sensor is Sensirion company's SHT11 digital type air temperature and humidity sensor.

In the experimental field, we choose the jonquil that grows well to place the sensor. Soil moisture sensors are buried in the ground near the root systems of plants, light sensors and air temperature and humidity sensors together with the ZigBee module are served as data acquisition node hanging above the plant. The electromagnetic valve is

connected with the router node driving circuit. The router is connected to the control machine through the RS485 serial port.

The router detects all nodes in the network, including data collection nodes and control nodes. The router receives the data sent by the nodes through the ZigBee wireless sensor network transceiver and uploads the data to the controller through the serial port. Controller uses fuzzy control algorithm to analyze the data and determines whether the need for irrigation. The decision transmits to the control nodes through the ZigBee wireless sensor network transceiver. The control nodes open or close the electromagnetic valve according to the command.

Jonquil's cultivation in the facility agriculture and environment experiment field of our school's water park is shown in the Fig.7.



**Fig. 7.** Jonquil's cultivation in our experiment field

## 6 Conclusions

This paper presents the technology of applying a ZigBee wireless sensor network to intelligent irrigation. With ZigBee chip CC2530 as the core, combined with the controller, the router, data acquisition nodes, the control nodes and irrigation device, the intelligent irrigation is realized through the wireless network.

The system has been carried out in the experimental field of our school and good communication effect has been obtained. Preliminary test shows that the system runs stably and reliably. Moreover, it can obtain the network node attributes accurately to control water saving irrigation. By further adjustment and testing, extension of this technique is not only conducive to the development of agriculture, but also is

significant to alleviate the shortage of water resources and improve the condition of agricultural water utilization.

Application of the ZigBee wireless sensor networks in the intelligent irrigation technology for digital agricultural information collection network is very valuable, which will promote the modernization of China's agriculture technology a lot.

## References

1. Chen, S.-H., Zheng, Z.-W., Sun, Z.-G., Wu, Q.-H.: A research on technique of linear networking based on Zigbee wireless sensor network. In: 2011 IEEE 3rd International Conference on Communication Software and Networks (ICCSN), May 27-29, pp. 263–266 (2011)
2. Xiao, J., Liu, X.: The research of E-AOMDVjr routing algorithm in ZigBee network. In: Control and Decision Conference (CCDC), May 23-25, pp. 2360–2365 (2011) (Chinese)
3. Deng, J., Deng, C., Xie, C., Quan, S.: Development of communication interface of AC charging for charging station based on ZigBee. In: 2011 International Conference on Consumer Electronics, Communications and Networks (CECNet), April 16-18, pp. 2130–2133 (2011)
4. Kang, M.-S., Ke, Y.-L., Li, J.-S.: Implementation of smart loading monitoring and control system with ZigBee wireless network. In: 2011 6th IEEE Conference on Industrial Electronics and Applications (ICIEA), June 21-23, pp. 907–912 (2011)
5. Dunkels, A., Alonso, J., Voigt, T., Ritter, H., Schiller, J.: Connecting Wireless Sensors with TCP/IP Networks. In: Langendoerfer, P., Liu, M., Matta, I., Tsoussidis, V. (eds.) WWIC 2004. LNCS, vol. 2957, pp. 143–152. Springer, Heidelberg (2004)
6. Hui, J.W., Culler, D.E.: Extending IP to low-power, wireless personal area networks. IEEE Internet Computing, 37–45 (July/August 2008)
7. Mao, Q.-J.: Design for ZigBee wireless sensors nodes used in environmental monitoring. Computer Knowledge and Technology 5(1), 232–234 (2009)
8. IEEE Computer Society, Part 5.4: Wireless Medium Access Control and Physical Layer Specifications for Low Rate Wireless Personal Area Networks (LR-WPANs) (October 1, 2003)
9. Elmas, C., Deperlioglu, O., Sayan, H.: Adaptive Fuzzy Logic Controller for DC-DC Converters. Expert Systems with Applications, 1541–1548
10. Li, L.-F., Liu, X.-Y., Chen, W.-F.: A Variable Universe Fuzzy Control Algorithm Based on Fuzzy Neural Network. Proceedings of the IEEE 7

# M/I Adaptation Layer Network Protocol for IoT Based on 6LoWPAN

Zhihong Qian, Yijun Wang, Xue Wang, and Shuang Zhu

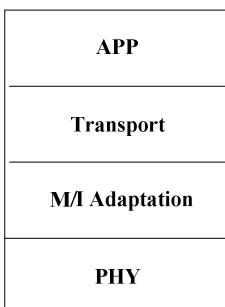
College of Communication Engineering, Jilin University, Nanhu Avenue,  
5372, 130012, Changchun, China  
{dr.qzh,wyjs-107}@163.com, yeti\_1019@yahoo.com.cn

**Abstract.** The four-layer network protocol architecture for Internet of Things is proposed based on 6LoWPAN, and the M/I adaptation layer, as the key of the protocol architecture, achieve the integration between underlying networks and Internet. MAC sublayer implement periodic listen/dormant mechanism and carry out unslotted CSMA/CA protocol in condition, and simultaneously fragmentation and reassembly of link frame header is used for link MAC packet. IP sublayer implements EHC scheme which could compress IPv6 global address header, and a IPv6 address autoconfiguration method is proposed in IP sublayer. The test results indicate that the network architecture is well for large-scale Internet of Things and related methods save the network energy and improve the system throughput.

**Keywords:** 6LoWPAN, Network Protocol, WSNs, Internet of Things.

## 1 Introduction

The Internet of Things (IoT) has two key ideas. One is that Internet is the core and foundation of IoT and the other is everything becomes terminal users that could communicate with each other. IoT has three primary application processes: (1) Identify the objects properties. Static property information is saved in RFID tag [1], and dynamic property information need to be detected by sensors all the time [2]. (2) Recognition equipments read the object information and transform into data format by the requirement of networks. (3) Object information is transferred into information processing platform by networks, which completes related application calculation [3].

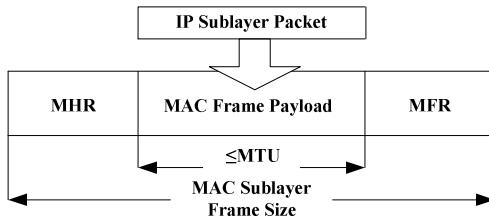


**Fig. 1.** IoT M/I adaptation protocol architecture

However, an important challenge for IoT is the integration between Internet and the underlying heterogeneous networks [4]. IEEE 802.15.4 is a short-distance wireless network communication standard. IPv6 is the dominant technology in the next generation Internet network layer. 6LoWPAN (IPv6 over Low power Wireless Personal Area Networks) defines how to carry IPv6 packets over IEEE 802.15.4 low power networks. As shown in Fig. 1, this paper defines M/I (MAC/IP) adaptation protocol architecture for IoT based on 6LoWPAN, and focuses on the design of M/I adaptation layer that embed IPv6 into IEEE 802.15.4.

## 2 Analysis of 6LoWPAN

The 6LoWPAN Work Group introduces an adaptation layer between data link and network layer, which realizes transmission of IPv6 Packets over IEEE 802.15.4 Networks. As we known, the data units on the Internet are transmitted by packets, so the IP packets become the data of MAC frame when it is sent to MAC layer. However, MAC protocol requires maximum length of data portion of the frame, MTU (Maximum Transfer Unit). Fig. 2 shows the relationship between MTU and data. Moreover, starting from a maximum physical layer packet size of 127 octets and a maximum frame overhead of 25, the resultant maximum frame size at the MAC layer is 102 octets. If starting Link-layer security, it imposes further overhead. This is obviously far below the minimum IPv6 packet size of 1280 octets. Thus, IEEE 802.15.4 MAC frame cannot encapsulate whole IPv6 packet [5-6].



**Fig. 2. MAC Frame Payload and MTU**

## 3 Design for M/I Adaptation Layer

This paper designs the M/I adaptation layer for satisfying the characteristics of the unified structure compared with 6LoWPAN. M/I layer provides an easy, flexible, unlink and best-effort delivered datagram service up to the application layer; while it completes control to the network topology, network routing and network building with no need to consider how the physical layer to realize a bit transmission in details.

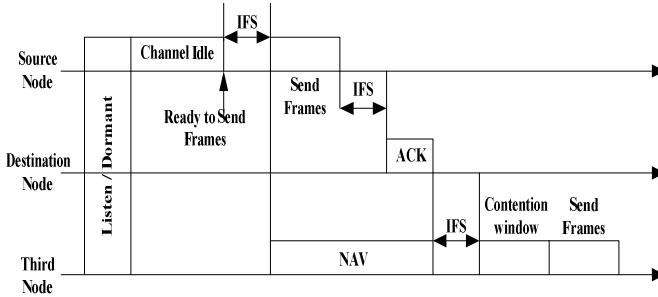
### 3.1 MAC Sublayer

Compared with traditional computer network, the IOT has characteristics of low power, low rate, low cost and large scale. Therefore, the traditional network MAC

protocol cannot directly applied to the IOT. We adopt a new CSMA/CA protocol as MAC sublayer access protocol of the IOT.

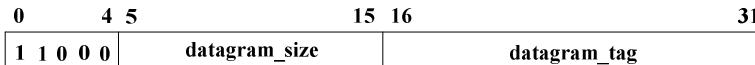
The basic mechanism describes as follow: nodes in the network use a periodic listening/dormant mechanism to reduce energy consumption. After the dormancy, a node will immediately change into dormant state if there is no activation events occurred that need the node to continue to work within the IFS (InterFrame Space).

When the node receives the signal of activation event, it will immediately switch to listening state, and then prepare to send date frame. Simultaneously it implements non-slotted CSMA/CA protocol conditionally. Periodic listen / dormant CSMA / CA protocol scheme is shown as Fig. 3.

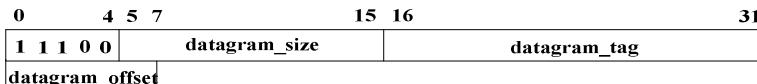


**Fig. 3.** Periodic listen / dormant CSMA / CA protocol scheme

If the payload submitted by IP sublayer protocol is larger than that of MAC sublayer MTU, the source node will split the payload, and needs to use the fragmentation header to provide reassembly information. The first and the subsequent fragmentation header structures are shown as Fig. 4 and Fig. 5.



**Fig. 4.** First Fragment



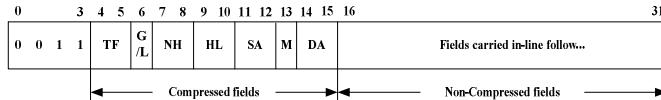
**Fig. 5.** Subsequent Fragments

## 3.2 IP Sublayer

### 3.2.1 IPv6 Header Compression

RFC4944[7] presents a header compression scheme of HC1, but it only carry out header compression against local link addresses, which loses the compression function for global addresses. Therefore, the paper proposes a wider-applied header file compression

scheme-EHC (Extensive Header Compression) based on HC1 scheme. Shown as Fig. 6, EHC structure is divided into compressed fields and non-compressed fields:



**Fig. 6.** EHC Header Compression Scheme

IP Sublayer could compress IPv6 global address header by embedding G/L(Global/Local) bit, HL(Hop Limit) bit and M(Multicast) bit in head compression encoding, and specific forms are as following:

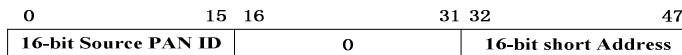
- G/L. 0: Global Address, 1: Link Local Address.
- HL. 00: Field Values is 1, 01: Field Values is 255, 10: the field is compressed, 11: the field is uncompressed.
- M. 0: Destination Address is unicast address, 1: Destination Address is multicast address.

### 3.2.2 IPv6 Address Autoconfiguration

#### a) Formation of Link Local Address

The equipment with IEEE 802.15.4 protocol has a long address of IEEE EUI-64 and a 16-bit short address as its MAC address.

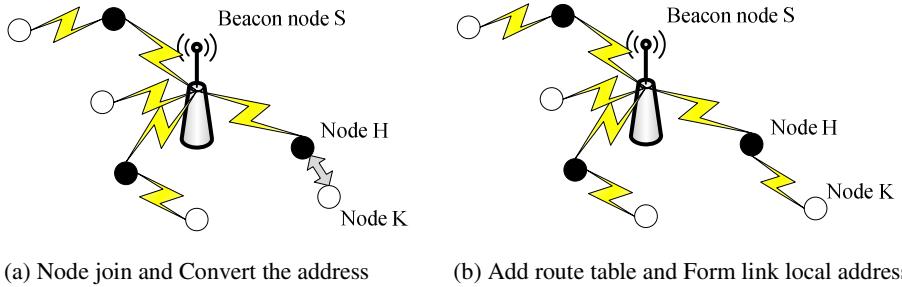
We use 16-bit short address to form link local address. For 16-bit short address, it will convert address into 48-bit standard MAC address through the method as following. Firstly, the far left 32-bit uses the PAN ID of source node, and then adds 16 zero behind it; secondly, this 32-bit field combines with the 16-bit short address of the source node, and then forms a 48-bit standard MAC address. The structure is shown as Fig. 7.



**Fig. 7.** 48-bit Standard MAC address

Then we use this 48-bit standard MAC address to obtain the interface ID of the auto-configured IPv6 address: insert a 16-bit encoded FFFE in the middle of above-mentioned 48-bit MAC address, and then convert to 64-bit interface ID of IPv6 address. The structure is shown as Fig. 8. Finally autoconfiguration of overall link local address can be completed, when we add link local address prefix FE80::/64 to the front of the interface ID, the diagram as shown in Fig. 9. When a new node K wants to join the network and communicates with the node H, it will obtain standard 48-bit MAC sublayer address by 16-bit short address, and then it convert to EUI-64 address of the interface identifier. After obtaining the 64-bit address of the interface identifier, report the 16-bit short address of the node K to the beacon node S by the node H, then the beacon node S adds this 16-bit short address to the routing table, furtherly add link local address prefix FE80::/64 to the front of the interface identifier.

64 bit	64 bit		
Prefix	Interface ID		
	FFFE(H)		
	24 bit	16 bit	24 bit

**Fig. 8.** 16-bit short addresses Transformation Format**Fig. 9.** Link Local Address**b) Formation of the Global Address**

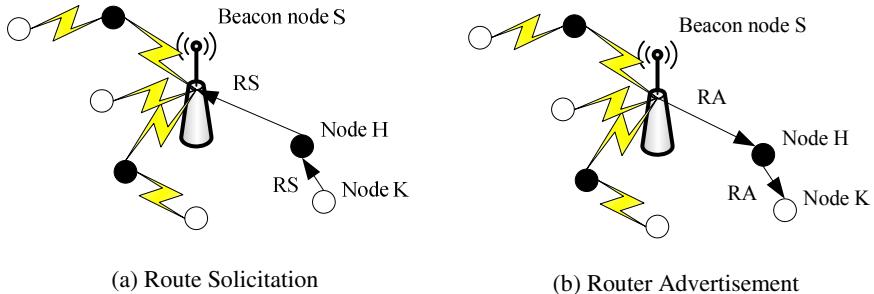
For 64-bit long address, its 64-bit extension identifier forms interface identifier (ID) of IPv6 address. The structure that is 64-bit long address converts to IPv6 address is shown as Fig. 10. The seventh bit in the first 24 manufacturer identification in IEEE EUI-64 is U bit, namely, unified/local (U/L), where U=1 expresses local administered address mode, U=0 expresses unified administered address mode; the eighth bit is I bit, namely, Unicast/Multicast (I/G), I=1 expresses multicast, I=0 expresses Unicast.

64 bit	64 bit	
Prefix	Interface ID	
	Manufacturer ID	Extension ID
	24 bit	40 bit

**Fig. 10.** 64-bit long addresses Transformation Format

As shown in Fig. 11, in order to obtain global routing prefix, the node K needs to send RS(Route Solicitation) to its adjacent beacon node S; the source address of the RS messages is the link local addresses of the node K, and the destination address is the Multicast address of the beacon node S. After receiving the RS message of the node K, the beacon node S will return a RA (Router Advertisement) to the node K, the source address of RA message is IPv6 address of the beacon node S, and the destination address is the link local address of the node K. Meanwhile, the option

field of this RA message will provide the global routing prefix. After obtaining the global routing prefix, a global IPv6 address is formed by the node K that combines with the interface identifier. So far, address autoconfiguration process of the sensor nodes is completed.



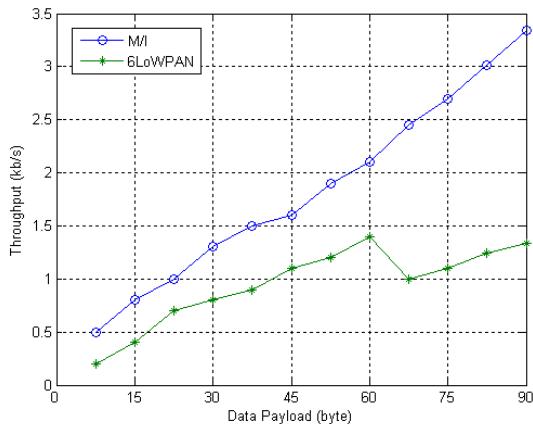
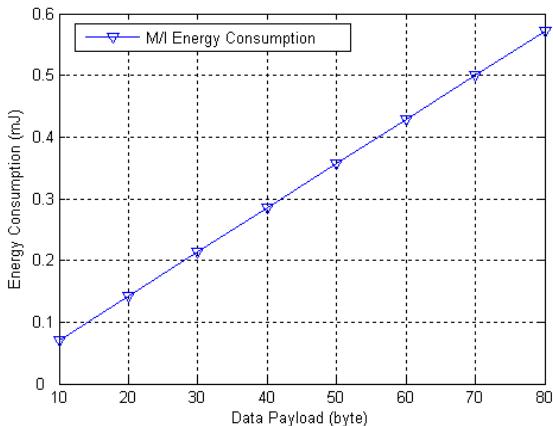
**Fig. 11.** Route Solicitation and Advertisement

## 4 Experiment

For the M/I layer application, we use Xi'an Huafan Technology HFZ-CC2430ZDK microcontrol model, with a single-chip 2.4 GHz RF transceiver (CC2430-F128 chipset). It also includes 128K Bytes ROM, 4K Bytes RAM. The PAN nodes are connected through the PPP interface to the Beacon node. We build implementation testbed over M/I stack. The devices support AODVjr and Cluster-Tree routing protocol.

The M/I Beacon node is the gateway that is connected to Internet. In addition, the Beacon node is coupled with the other nodes within the network by 16-bit short address. After initialization its address is 0. All the nodes form the network based on the Beacon node.

We use testbed described above to evaluate the network throughput compared with 6LoWPAN at first. Figure 12 shows the average system throughput under the two cases along with the increasing data payload. From the Fig. 12, M/I has improved about 30% in throughput compared with 6LoWPAN, which is the important advantage of this protocol. Secondly, the experiment testifies the reliability of the M/I protocol from the perspective of energy consumption. From Fig. 13, the energy consumption increases linearly with the increase of data payload bytes, but it is always maintained at the level of 0.1mJ. Even though transmission payload achieves maximum bytes for 127, the energy consumption still remains within the 1mJ; while the typical wireless communication protocol including 6LoWPAN needs to consume several millijoules in variety when it transmits one byte. Therefore, the improved light network protocol reduces the energy consumption, which lays the foundation for large-scale popularization of the IoT.

**Fig. 12.** Throughput Analysis**Fig. 13.** Network Protocol Energy Analysis

**Acknowledgments.** This work is supported by the National Natural Science Foundation of China (No.60940010, No.61071073) and the Doctoral Fund of Ministry of Education of China (No. 20090061110043).

## References

1. Wang, X., Qian, Z.H., Hu, Z.C., Li, Y.N.: Research on RFID anti-collision algorithms based on binary tree. *Journal on Communications* 31, 49–57 (2010)
2. Kranz, M., Holleis, P., Schmidit, A.: Embedded Interaction: Interacting with the Internet of Things. *J. IEEE Internet Computing* 14, 46–53 (2010)

3. Antonio, J., Miguel, A.Z., Antonio, F.G.: An architecture based on Internet of Things to support mobility and security in medical environments. In: 7th IEEE Consumer Communications and Networking Conference, pp. 1–5. IEEE Press, New York (2010)
4. Zhu, Q., Wang, R.C., Chen, Q.: IoT gateway: Bridging wireless sensor networks into Internet of Things. In: IEEE/IFIP International Conference on Embedded and Ubiquitous Computing, pp. 347–352. IEEE Press, New York (2010)
5. Mazzer, Y., Tourancheau, B.: Comparisons of 6LoWPAN Implementations on Wireless Sensor Networks. In: 3th International Conference on Sensor Technologies and Applications, pp. 689–692. IEEE Press, New York (2009)
6. Wang, Y.J., Qian, Z.H., Wang, X., Sun, D.Y.: Addressing Scheme for Internet of Things Based on 6LoWPAN. *Journal of Electronics & Information Technology* 34, 763–769 (2012)
7. Montenegro, G., Kushalnagar, N., Hui, J.: Transmission of IPv6 Packets over IEEE802.15.4 Networks. S. IETF RFC 4944 (2007)
8. Zhu, H.B., Yang, L.X., Zhu, Q.: Survey on the Internet of Things. *Journal of Nanjing University of Posts and Telecommunications (Natural Science)* 31, 1–9 (2011)

# A Novel Method to Improve Gain and Tune Impedance of RFID Tag Patch Antenna

Chuncheng Kong and Jun Hu

Center for Optical and Electromagnetic, Zhejiang Provincial Key Lab for Sensing Technologies, State Key Lab of MOI, Zhejiang University, China  
kunchong0120@163.com, hujun@zju.edu.cn

**Abstract.** The effect of ground plane's extension of RFID tag patch antenna on its gain and impedance is discussed. The ground plane's extension at radiating slot can effectively improve gain and will visibly change impedance together with that at non-radiating slot. A general variation law of antenna performance with ground extension is proposed. It's convenient and low-cost to get higher gain and desired impedance through proper ground extensions along both directions rather than designing ground plane meaninglessly. This paper takes an existing RFID patch antenna as an example to verify the proposed theory is effective.

**Keywords:** patch antenna, ground plane, radiating slot, non-radiating slot, better performance.

## 1 Introduction

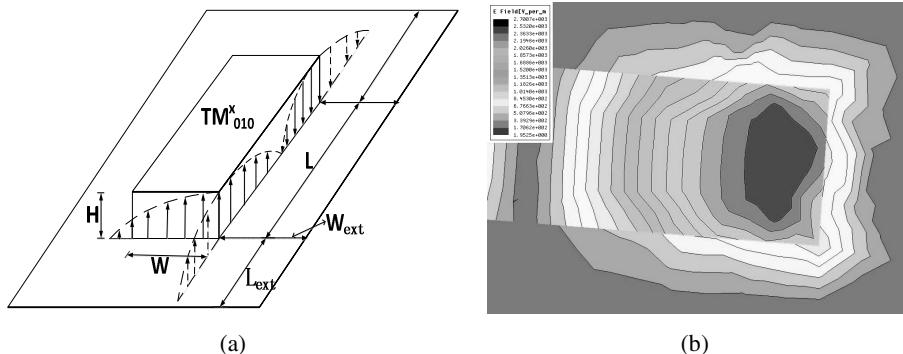
Radio frequency identification (RFID) technology has gained fast development in many fields such as supply chain management, service industries and sensor network recent years. As an important part of RFID system, gain and impedance of tag antenna have a big influence on RFID system's performance. It's significant to find a general way to improve antenna gain and get appropriate impedance on different attached surfaces. Many tag patch antennas have been proposed to acquire better performance on different surfaces especially metallic surface [1-3] in comparison to dipole-like antennas. However, most of these antennas are based on some special complexly geometric designs or ground structures such as via holes which add cost and may have unexpected result of antenna performance. It's necessary to find a general way with low cost and slight structure change for tag patch antennas to improve performance.

In this paper, the effect of ground plane's extension of patch antenna is researched. Through extending ground plane at radiating slot and non-radiating slot, gain and impedance of antenna can change in wide range with predictable results. An existing RFID tag patch antenna prototype [4] is discussed for theoretical study and model simulation.

## 2 Theoretical Research

As we all know, according to cavity theory patch antenna has two radiating slots and two non-radiating slots. The radiating slots can generate effective radiation while

non-radiating slots' radiation will be counteracted each other. For most existing RFID tag patch antenna designs, the ground planes are general in the same size with substrates or just big enough. But it's not a wise choice to obtain better antenna performance. Proper extension of ground plane can improve gain effectively and make impedance comparatively stable with different attached objects.



**Fig. 1.** (a) The field configuration (modes) of  $TM^x_{010}$  for patch antenna. (b) part of E-Field distribution of the ground plane

Using cavity theory, the equivalent magnetic current density  $\mathbf{M}_s$  around the side periphery of the patch generates radiation into free-space. The presence of the ground plane can be taken into account by image theory which will double  $\mathbf{M}_s$  [5]. But actually the ground plane cannot be infinite which means  $\mathbf{M}_s < 2\mathbf{n} \times \mathbf{E}$ . The longer ground plane's extension is, the notable image effect is. In other words, there is more energy gathered on field modes equivalently. So by extending ground plane at radiating slot, antenna gain will be improved because radiation field mode gets more energy. In other hand, antenna gain will be reduced since non-radiation field mode gets more energy. The gain cannot be improved limitlessly as the ground plane at radiating slot become bigger since more field modes will be generated making energy dispersive and counteracted.

There are many field modes in patch antenna and this paper takes only  $TM^x_{010}$  mode into consideration since it is the dominant field mode with lowest resonance. Fig.1(a) shows the  $TM^x_{010}$  field mode of patch antenna.  $L$ ,  $W$  and  $H$  represent respectively dimension of radiating slots, non-radiating slots and substrate height with  $L>W>>H$ . The  $L_{ext}$  and  $W_{ext}$  are respectively the size of ground plane's extension at radiating slot and non-radiating slot.

When the mode transmits into air, the equivalent wavelength  $\lambda_{air}$  will be shorten due to the influence of ground plane which can be calculated by

$$\lambda_{air} = \frac{\lambda_r}{\sqrt{\epsilon_r}} = \frac{2L_{eq}}{\sqrt{\epsilon_r}} \quad (1)$$

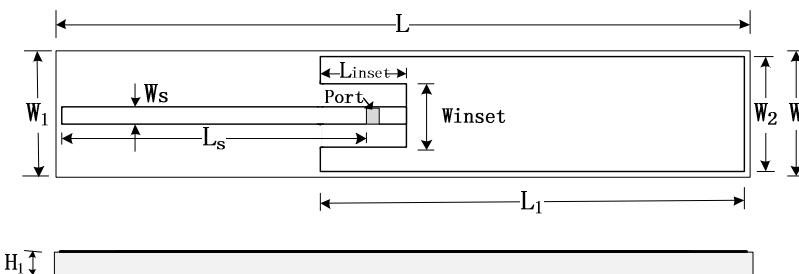
Where  $\epsilon_r$  is the permittivity of substrate,  $\lambda_r$  is the  $TM^x_{010}$  resonant wavelength and  $L_{eq}$  is the equivalent patch length [5]. This formula is first proposed for calculating the equivalent wavelength in air above ground plane.

As shown in Fig.1(a), when  $L_{ext} < \lambda_{air}/4$ , the electricity vector direction of fringing field beyond the substrate is still unchanged, so gain of antenna is monotonically increasing and impedance is decreasing in this range. If  $L_{ext} > \lambda_{air}/4$ , the fringing field will have opposite electricity vector direction which makes field strength reduced and impedance increased. In consideration of other field modes, the gain and impedance will oscillate when  $L_{ext}$  becomes bigger than  $\lambda_{air}/4$ .

The variation tendency of gain and impedance on  $W_{ext}$  can be explained in the same way. When  $W_{ext} < \lambda_{air}/4 - W/2$ , gain of antenna is monotonically decreasing and impedance is increasing. When  $W_{ext} > \lambda_{air}/4 - W/2$ , the gain and impedance will oscillate.

### 3 Simulated Results and Analysis

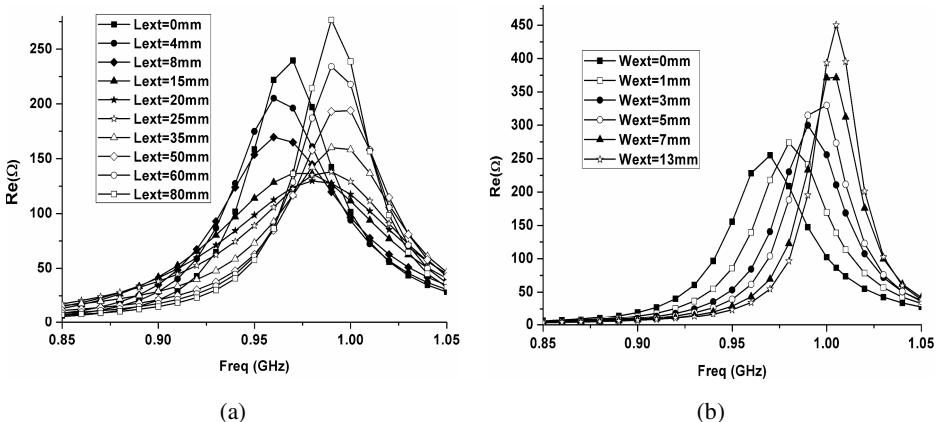
An existing RFID tag patch antenna [4] is simulated with Ansoft HFSS V12 for validating theory model, as shown in Fig.2. The resonance frequency simulated by HFSS is 1.01GHz so the equivalent patch length  $L_{eq}$  is 69.2mm calculated by cavity theory [5] less than  $L_1$  (74.5 mm).



**Fig. 2.** RFID tag patch antenna prototype used for simulation

The simulated E-field figure of the ground plane is shown in Fig.1 (b). The E-field is discontinuous at the interface of ground plane under the substrate and the extensional part. That means the electric field is not spreading directly from inside out on the ground plane. What's more, the E-field of extensional ground plane is compressed with smaller range than that under substrate, which confirms our theoretical assumption that the effective wavelength will be shorten beyond the substrate.

Since impedance and gain have the same monotonic change tendency with  $L_{ext}$  or  $W_{ext}$ , only the relationship of variation between resistive component of impedance and  $L_{ext}$  and  $W_{ext}$  are shown in Fig.3.  $L_{ext}$  is just changed at right side of antenna and  $W_{ext}$  at both sides. The length between patch and substrate boundary which is 1mm is all not included in  $L_{ext}$  and  $W_{ext}$ . The maximal impedance and resonance are both decreasing monotonously until  $L_{ext}$  reach 15mm which is in agreement with theory value of 15.13 mm with  $\epsilon_r=4.6$ . Meanwhile, the theoretical first inflection point of  $W_{ext}$  is 5.13mm while the simulated value is 5mm.



**Fig. 3.** Resistive component of antenna impedance changed with different (a)  $L_{\text{ext}}$  (b)  $W_{\text{ext}}$

## 4 Conclusions

This paper shows that  $L_{\text{ext}}$  (length of ground plane's extension at radiating slot) can efficiently improve gain of RFID tag patch antenna and can also tune impedance in a large range with  $W_{\text{ext}}$  (length of ground plane's extension at non-radiating slot). The proposed theory also points out the variation tendency of antenna gain and impedance on  $L_{\text{ext}}$  and  $W_{\text{ext}}$ . Since a general law is present, it's convenient to design proper  $L_{\text{ext}}$  and  $W_{\text{ext}}$  for impedance matching for different objects which RFID tag is attached on.

## References

1. Dacuña, J.: Low-Profile Patch Antenna for RF Identification Applications. *IEEE T. Micro Theory* 57(5) (2009)
2. Björninen, T.: Long Range Metal Mountable Tag Antenna for Passive UHF RFID Systems. In: 2011 IEEE International Conference on RFID-Technologies and Applications (RFID-TA), pp. 202–206 (2011)
3. Wu, J., Li, J.: Miniaturized Dual-Band patch Antenna Mounted on Metallic Plates for RFID Passive Tag. In: 2011 International Conference on Control, Automation and Systems Engineering (CASE), pp. 1–4 (2011)
4. Mo, L., Qin, C.: Planar UHF RFID Tag Antenna With Open Stub Feed for Metallic Objects. *IEEE T. Antennas Propag.* 58(9) (2010)
5. Balanis, C.A.: *Antenna Theory Analysis and Design*, 3rd edn. A John Wiley & Sons, Inc., Publication (2005)

# Improved Accuracy of RFID Localization Assisted by Output Power Adjustment of the Reader<sup>\*</sup>

Xiaoyin Li, Lianshan Yan, Wei Pan, Bin Luo, and Q.F. Guo

School of Information Science and Technology,  
Southwest Jiaotong University, Chengdu, Sichuan, China, 610031  
brosvip@126.com

**Abstract.** In this paper, we present a simple and effective radio frequency identification (RFID) localization method with good accuracy using passive UHF RFID tags. To achieve the accurate location, a path loss model is proposed to study the power of the reader as a function of read range. The function is analyzed based on an empirical model that often used in indoor environments in theory and demonstrated in experiments. Based on the function, a circle localization using multiple RFID antennas is proposed. Experiment and simulation result show that the localization accuracy could reach tens of centimeters by adjusting the output power of the RFID reader using two or three antennas.

**Keywords:** Radio frequency identification (RFID), localization, read range, intersection.

## 1 Introduction

Many applications need to know the physical location of objects, so automatic location-sensing systems have become very popular in recent years [1]-[3]. On the other hand, Radio-frequency identification (RFID) technique is being such a hot topic due to the emerging Internet of Things (IOT) [4]-[5], and the localization based on RFID has received considerable attentions in recently years [6]-[7]. According to the type of the target, RFID-based localization systems can be divided into two classes: reader localization and tag localization [8]. Many UHF RFID based localization techniques presented are reader-based ones, and the accuracy is limited by the separation of reference tags (e.g. 30 cm in [9]), therefore, a large number of reference tags are required, and only the target equipped with bulky RFID Reader can be located. Such scheme is not suitable for the location of numbers of targets and the localization region is also limited. On the other hand, some tag localizations are demonstrated, but the accuracy is poor, e.g. only 1.6m in [10].

In this paper, we propose a novel tag localization method with improved accuracy based on adjusting the power of UHF RFID reader with multiple antennas.

---

\* The research is supported by the National Natural Science Foundation of China (No. 60972003).

Experiment demonstration using two antennas shows the accuracy varying between 9 and 260 cm, with an average value of  $\sim 57$  cm. We further compare the performance using two and three antennas through simulating 600 random targets within a 4m x 4m region. Statistical results of the location accuracy in terms of 95% probability indicate that the accuracy is improved from 140 cm to 50 cm using three antennas. The key advantages of our approach include: (i) there is no reference tag needed thus to reduce the complexity of the system; (ii) it can be potentially used for multiple targets' location without modifying the setup (i.e. tag localization).

## 2 Concept and Experimental Setup

Due to the complexity of factors that have influence in RFID read range, the path loss model in free space is not suitable for the indoor environment. An empirical model based on a two-slope model which is often used in indoor environments is given in [11]. As for passive RFID, R0 is longer than the maximum read range, the empirical model can be simplified by taking into account only the first path loss term, and this is [11-13]:

$$L_p(dB) = -20 \log\left(\frac{\lambda}{4\pi}\right) + n_1 10 \log(r) + L_{obs}(dB) \quad (1)$$

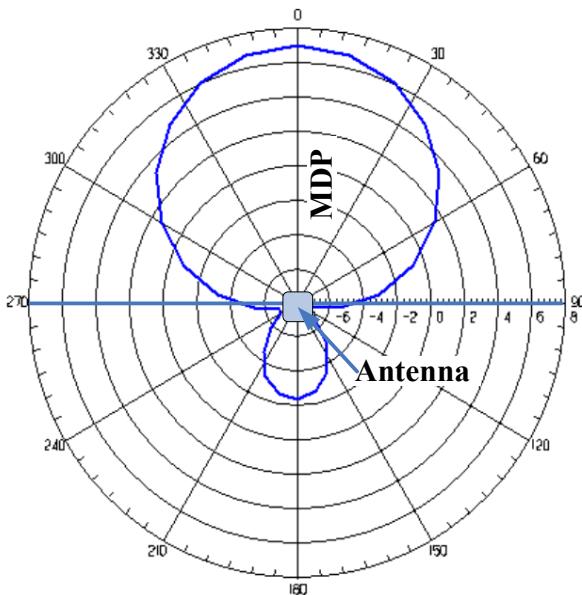
The power received in tag ( $P_{r,tag}$ ) is given by a modification of Friis transmission equation in (2) [11]:

$$\begin{aligned} P_{r,tag}(dBm) &= P_{reader}(dBm) + G_{reader}(dB) + G_{tag}(dB) \\ &\quad + 10 \log(1 - |\rho|^2) + \Delta G(dB) \\ &\quad - L_{sys}(dB) - L_p(dB) \end{aligned} \quad (2)$$

From equation (1) and (2), the relationship between the power of the reader and the read range can be deduced as (3):

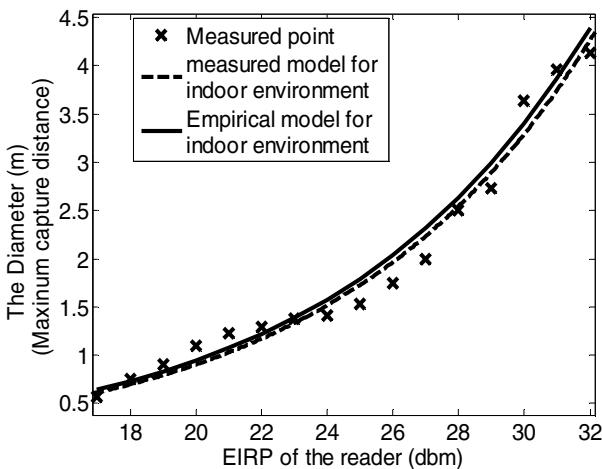
$$\begin{aligned} P_{reader}(dBm) &= P_{r,tag}(dBm) - G_{reader}(dB) - G_{tag}(dB) \\ &\quad - 10 \log(1 - |\rho|^2) - \Delta G(dB) \\ &\quad + L_{sys}(dB) + L_p(dB) \end{aligned} \quad (3)$$

The RFID antenna (XCAF-12L) we used is circular polarized, and its radiation pattern is shown in Fig.1. The read range (the maximum distance at which the tag can be read by the reader) of the reader can be determined by the output power [14]. Instead of using reference tags, we can first obtain the relationship between the read range and the power of the reader, and subsequently scan or locate the target through power adjustment (equivalent to scanning the capture range).



**Fig. 1.** The radiation pattern of the antenna

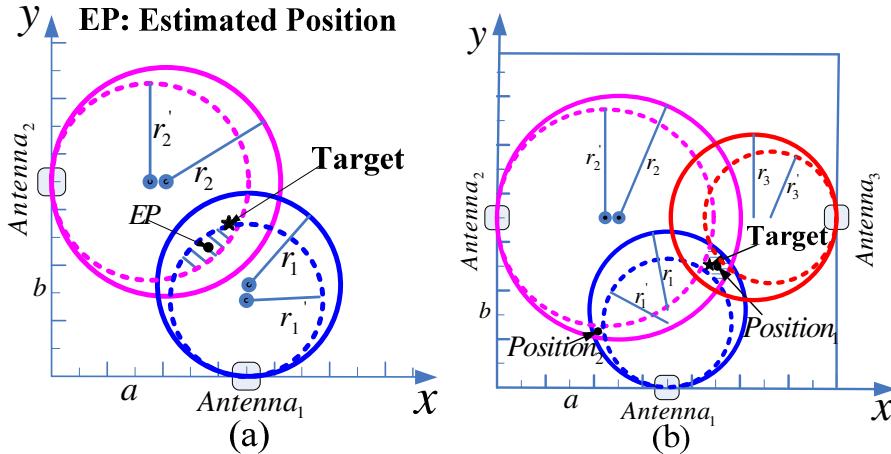
Experimental measurements are first conducted to obtain the function between the maximum read range and the power of the reader, and compared with the theoretical model (with the path loss exponent is 1.8 for indoor LOS [13]) as shown in Fig.2.



**Fig. 2.** The relationship between read range and the power of the reader

The proposed configurations are shown in Figs .3(a) and 3(b) using two and three antennas, respectively. From the above relationship and use Fig. 3(a) as an example (i.e. two antennas), we can adjust the power of the antenna in x axis until it reaches

the minimum power  $P_x$  for detecting the target tag. In our experiments, the EIRP is automatically adjusted from 17 to 32dBm with 0.5-dB as the step (more precise step may not be necessary as it is hard to find the localization difference for small power variations). Similarly, we can get the minimum power  $P_y$  for detecting the target tag in y axis. After that, we can substitute  $P_x$  and  $P_y$  into the function to get corresponding  $D(p)$  values, i.e.  $D_x$  and  $D_y$ .



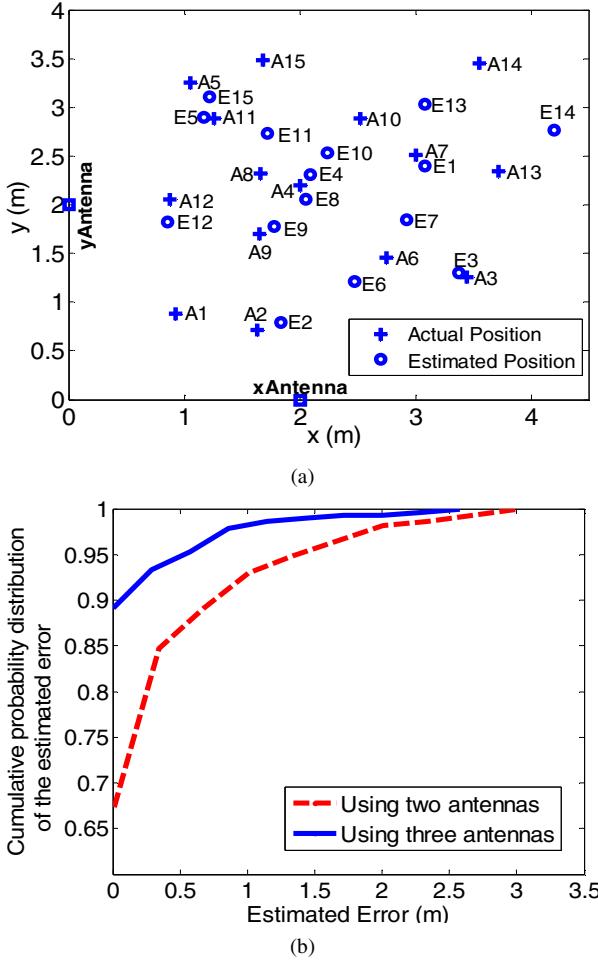
**Fig. 3.** The principle of localization (a) using two antennas (only one intersection shown) and (b) using three antennas

Now we can draw the patterns of the two antennas at  $P_x$  and  $P_y$ , as well as  $P_x-0.5$  (or  $P_x+0.5$ ) and  $P_y-0.5$  (or  $P_y+0.5$ ) (power all in dBm). The four patterns construct one or two intersections, in which the target is located. We calculate the cavity of intersections and use it (or the center of two) as the estimated position of the target (Fig.3). The location of the target is restricted by Eq. (4):

$$\text{s.t. } \begin{cases} (x-a)^2 + y^2 \leq r_1^2 \\ (x-a)^2 + y^2 \geq r_1'^2 \\ x^2 + (y-b)^2 \leq r_2^2 \\ x^2 + (y-b)^2 \geq r_2'^2 \end{cases} \quad (4)$$

### 3 Results and Discuss

To verify the feasibility of the approach, we conduct the localization experiment with the target tag randomly scattered at 15 positions. Fig. 4 (a) illustrates the localization accuracy for these positions (“A” stands for actual position and “E” stands for the estimated one). The estimation error varies from 9 cm to 260 cm, (most of the estimated error within 100cm, and only one is up to 260cm), with an average value of ~ 57 cm.



**Fig. 4.** Results of localization (a) experimental results with ten randomly scattered positions; (b) simulation results of accuracy comparison (statistical probability vs. estimation error) between two and three antennas with 600 randomly distributed target tags

A major error contribution comes from the fact that there are maybe two intersection areas (Fig. 3a is only a special case with one intersection while Fig. 3b is a more general case), but only one calculated cavity is valid (i.e. close to the actual position) and could not be determined easily. Therefore, we add another antenna (Antenna 3 in Fig. 3b) to more effectively determine the target location (due to limitations of automatic antennas control, this part is not done experimentally). For the evaluation of three antennas, we perform statistical simulation using MATLAB with 600 random target tags within an area of 4m \* 4m. The accuracy comparison is shown in Fig. 4(b). In terms of 95% probability, the accuracy is significantly improved from ~140cm to ~50cm using two and three antennas, respectively. 0.5-dB step is used in the simulation as well. Note that the accuracy may also be affected by the environmental perturbations (e.g. EM reflections, space size/shape, etc.).

## 4 Conclusion

We proposed a RFID-based localization scheme with significantly improved accuracy especially when three antennas are deployed. Through adjusting the power of the reader (or antennas), the localization accuracy can be within tens of centimeters or even better. The localization scheme can be a potential candidate for low-cost localization applications in future IOT systems.

## References

1. Liu, H., Darabi, H., Banerjee, P., Liu, J.: Survey of wireless indoor positioning techniques and systems. *IEEE Trans. Syst., Man, Cybern.–Part C: Appl. Rev.* 37(6), 1067–1080 (2007)
2. Viikari, V., Pursula, P., Jaakkola, K.: Ranging of UHF RFID Tag Using Stepped Frequency Read-Out. *IEEE Sensors J.* 10(9), 1535–1539 (2010)
3. Ni, L.M., Liu, Y., Lau, Y.C., Patil, A.P.: LANDMARC: Indoor location sensing using active RFID. In: Proc. IEEE PerCom, pp. 407–415 (2003)
4. <http://www.itu.int/osp/spu/publications/internetofthings/>
5. Yan, L., Zhang, Y., Yang, L.T., Ning, H.: The Internet of Things. Aurebach Publications (2008)
6. Park, S., Hashimoto, S.: Autonomous Mobile Robot Navigation Using Passive RFID in Indoor Environment. *IEEE Trans. Industrial Electronics* 56(7), 2366–2373 (2009)
7. Saab, S., Nakad, S.: A Standalone RFID Indoor Positioning System Using Passive Tags. *IEEE Trans. Industrial Electronics* 58(5), 1961–1970 (2011)
8. Papastolou, A., Chaouchi, H.: RFID-assisted indoor localization and the impact of interference on its performance. *J. Netw. Comput. Appl.* 34(3), 902–913 (2011)
9. Pathanawongthum, N., Chemtanomwong, P.: RFID based Localization Techniques for Indoor Environment. In: Int. Conf. Advanced Comm. Technol., vol. 2, pp. 1418–1421 (2010)
10. Xiong, T.W., Liu, J.J., Yang, Y.Q., Tan, X., Min, H.: Design and implementation of a passive UHF RFID-based Real Time Location System. In: Int.l Symp. VLSI Design Automation & Test (VLSI-DAT), pp. 95–98 (2010)
11. Lazaro, A., Girbau, D., Salinas, D.: Radio Link Budgets for UHF RFID on Multipath Environments. *IEEE Trans. Antennas Propagat.* 57(4), 1241–1251 (2009)
12. Wang, H.G., Pei, C.X., Zhu, C.H.: A Link Analysis for Passive UHF RFID System in LOS Indoor Environment. In: Int. Conf. Wireless Communications, Networking and Mobile Computing, pp. 1–7 (2008)
13. Duangsawan, S., Promwong, S., Sukutamatanti, N.: Measurement and Modeling of RFID Propagation Channel with in an Indoor Environment. In: Int. Conf. Advanced Computer Theory and Engineering, pp. 393–397 (2008)
14. Chen, Z.N., Qing, X., Chung, H.L.: A Universal UHF RFID Reader Antenna. *IEEE Trans. Microwave Theory Tech.* 57(5), 1275–1282 (2009)

# A Two-Layer Duplicate Filtering Approach for RFID Data Streams

Wen Jiang, Yongli Wang, and Gongxuan Zhang

School of Computer Science and Technology  
Nanjing University of Science and Technology, NJUST  
Nanjing, China  
jiangwennjust@gmail.com

**Abstract.** Duplicates in RFID data stream must be filtered before sent to applications. Existing methods only considered detecting the duplicates in the server side, thus wasting lots of bandwidth to transfer the duplicate data from the reader to the server. This paper first identifies the two sources of duplicates, namely local duplicates and global duplicates. Then we propose a two-layer approach to filter those duplicates. In the reader side, we devise a method based on Bloom Filters to filter local duplicates. This greatly reduces the bandwidth requirement to transfer those duplicates. In the server side, we devise Distributed Time Bloom Filters to filter global duplicates. Experimental results show that our approach greatly reduced the network bandwidth required to transfer the data and had linear scalability in processing rate. Our approach are most suitable to applications which generate lots of local duplicates and the bandwidth is a scarce resource.

**Keywords:** RFID Data Streams, Duplicate Filtering, Bloom Filters.

## 1 Introduction

RFID based systems consist of tags, which are attached to physical objects to track their movements, and readers, which can read the information in tags within detection region without contact. The low price of tags and the way that readers detect tags make RFID systems desirable for many applications, such as supply chain management [1], postal package tracking [2], etc. Many RFID systems have been deployed in the last few years. However, there are some technological challenges that limit the widespread adoption of RFID systems. The most important one is RFID data duplication [3]. Because RFID readers can detect tags without line of sight, there are many redundant data in raw RFID data streams. It's necessary to filter the duplicates before sending to the application; otherwise the application may go wrong, such as the 'COUNT' operator in SQL queries.

To eliminate duplicates in RFID data streams efficiently, some approximate approach based on Bloom Filters [4] have been proposed. Metwally et al. [5] use Bloom Filters to detect duplicates in click streams. However, since Bloom Filter doesn't support deletion operation, the Bloom Filter will soon be filled as data is

continually inserted and become useless. To solve the above problem, Deng and Rafiei [6] propose Stable Bloom Filters, which extend the original Bloom Filters to support elements deletion. Before inserting an element, it randomly selects some positions to decrease their counting values to make room for the new element. However, this method generates both false positive and false negative errors. Wang et al. [7] use Bloom Filters to filter duplicates over distributed data streams. They use a separate Bloom Filter for each data stream, and extend it to filter global duplicates by sharing the Bloom Filters. However, bandwidth is wasted by the Bloom Filter sharing process. Lee et al. [8] propose Time Bloom Filter to detect duplicate in RFID data streams, which extends original Bloom Filter to support sliding windows. However, they only consider filtering duplicates in the server side, so lots of bandwidth will be wasted for transferring the duplicates.

Our work is mainly motivated by [8]. The main difference is as follows. First, in order to reduce the network bandwidth requirement, we propose using a two layer approach to eliminate the local duplicates in the “network edge”, then detecting the global duplicates in the server side. Second, because all the RFID data will be sent to the server, the server will soon become overwhelmed, so the method in [8] doesn’t scale well. We propose distributing the Time Bloom Filters to several computing nodes to enhance the scalability.

The rest of this paper is organized as follows. Section 2 gives formal definitions about the duplicate data in RFID data streams, and identifies the two sources of duplicates. Section 3 presents our two-layer approach for filtering the duplicates, and gives a detailed description of the implementations. Section 4 demonstrates the effectiveness of our approach by carrying out experiments on a synthetic data set. And Section 5 concludes this paper.

## 2 Problem Definition

In this section, we give formal definitions about the RFID duplicate problem we’re considering, and identify the two sources of duplicates.

**Definition 1. (RFID Data Streams).** An RFID data stream is a sequence of items  $\langle x_1, x_2, \dots \rangle$ , where each item  $x_i$  is a triple (ReaderID, TagID, Timestamp). The “ReaderID” is the identification number for readers. The “TagID” is the identification number for tags. The “Timestamp” records the time when the reader detects the tag.

**Definition 2. (Duplicate).** In a RFID data stream, item  $x$  is considered as a duplicate if there exists item  $y$ , such that  $x.\text{TagID} = y.\text{TagID}$  and  $x.\text{Timestamp} - y.\text{Timestamp} < \tau$ , where  $\tau$  is application specific positive value.

Duplicates come from two sources. When a tag stays in the detection region of a reader, or moves too slowly, the tag might be read several times by the same reader.

**Definition 3. (Local Duplicate).** In a RFID data stream, item  $x$  is consider as a local duplicate if there exists item  $y$ , such that  $x.\text{TagID} = y.\text{TagID}$ ,  $x.\text{ReaderID} = y.\text{ReaderID}$ , and  $x.\text{Timestamp} - y.\text{Timestamp} < \tau$ , where  $\tau$  is application specific positive value.

When many tags move simultaneously through one detection region, a single reader may not be able to detect all of them. So in general, several readers are deployed to monitor a single location [9] to prevent missing readings. But multiple readers may read the same tag at the same time.

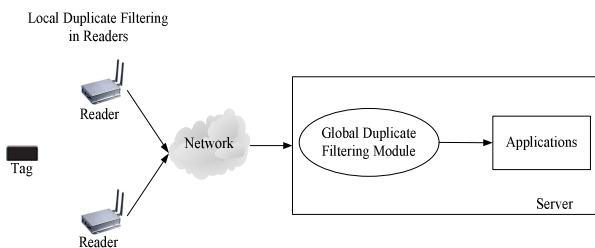
**Definition 4. (Global Duplicate).** In a RFID data stream, item  $x$  is consider as a global duplicate if there exists item  $y$ , such that  $x.\text{TagID} = y.\text{TagID}$ ,  $x.\text{ReaderID} \neq y.\text{ReaderID}$ , and  $x.\text{Timestamp} - y.\text{Timestamp} < \tau$ , where  $\tau$  is an application specific positive value.

### 3 Two-Layer Filtering Approach and Implementations

In this section, we will propose a two-layer approach for filtering duplicates in RFID data streams, and give a detailed description of the implementations in both layers.

#### 3.1 Two-Layer Filtering Approach

In order to reduce network bandwidth, it's important to filter the duplicates as early as possible. In this paper, we propose a two-layer approach to filter the duplicates, as shown in Fig.1. In each reader, we add a module to filter the local duplicates, which come from the same reader consistently reading the same tag. By detecting the local duplicates in the network edge, we can reduce the network bandwidth needed to transfer those duplicates. Then in the server side, we add a module to filter the global duplicates. In the following sections, we will give a detailed descriptions about how to filtering duplicates in both layers.



**Fig. 1.** Two-Layer filtering approach

#### 3.2 Preliminary: Bloom Filters

Bloom Filter was introduced by Burton Bloom in the 1970s [4], which is very space-efficient randomized data structure for representing set. Bloom Filter uses a bit array to represent a set. Suppose the array contains  $m$  bits, and the set  $S = \{x_1, x_2, x_3, \dots, x_n\}$  contains  $n$  elements. Then given an element  $x_i$ , Bloom Filter uses  $k$  independent hash functions to map this element into  $k$  positions in the bit array. If all bits in these positions are 1, then it reports  $x_i$  is in the set, otherwise, it reports  $x_i$  is not in the set. To insert element  $x_i$  in the set, we simply set all the bits in the

corresponding hashed positions as 1. Because different hash functions may map different element into the same position, Bloom Filter may generate false positive errors. According to [10], the false positive error rate is approximately  $(1 - e^{-kn/m})^k$ , and it reaches the minimum  $(0.6185)^{m/n}$  when  $k = (\ln 2)m / n$ . Thus by choosing  $m$  appropriately, we can make this error rate sufficiently low.

### 3.3 Filtering Local Duplicates

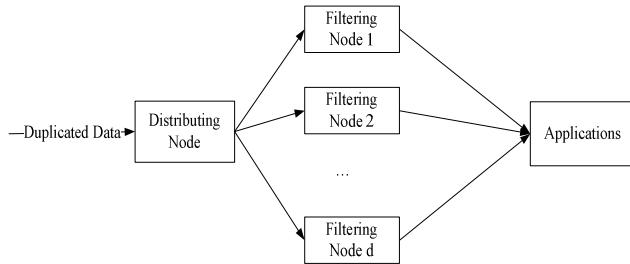
In this section, we devise an improved duplicate filtering algorithm based on Bloom Filters. The original Bloom Filter doesn't support deletion. According to the Definition 2, duplicates are related to a time threshold  $\tau$ . So in order to support these operations, we extend the original Bloom Filter as follows. Instead of using a single bit in each position of the array, we use  $t$  bits to represent the time to live for this element, where  $t = \lceil \log_2 \tau \rceil$ . We call this improved Bloom Filter TTL-BL in short. Initially, all the positions are set to 0. Then at each time unit, all the non-zero positions are decreased by one. The duplicate filtering process is shown in Algorithm 1 in Java language. In line 1 to 2, we compute the  $k$  hash positions and store in  $p_i$ . In line 3 to line 7, we check these  $k$  positions in the Bloom Filter to see whether this data appeared  $\tau$  time unit before. Then in line 8 to 9, we set these positions in the Bloom Filter to  $\tau$  to insert this data. Finally in line 10, we return the result.

```
Algorithm 1. TTL-BL(RFID_Data x)
Input: x is triple (ReaderID, TagID, Timestamp)
Output: true or false (whether x is a duplicate)
1:   for(i = 0; i < k; i++)
2:     pi = hashi(x.TagID); // hashi: ith hash function
3:   boolean duplicate = true;
4:   for(i = 0; i < k; i++)
5:     if(BL[pi] == 0) then //BL: array of time units
6:       duplicate = false;
7:       break;
8:   for(i = 0; i < k; i++)
9:     BL[pi] = τ;
10:  return duplicate;
```

The time complexity of Algorithm 1 is  $O(k)$ , where  $k$  is the number of hash functions. In hardware implementation, the  $k$  hash function can run concurrently.

### 3.4 Filtering Global Duplicates

The filtering module in the readers can only filter local duplicates. In this section, we give a description of the global duplicates filtering module in the server. Lee and Chung [8] propose Time Bloom Filters to detect duplicates in the server side. However, since all the data will flow through this module in the server, it will soon become the bottleneck. Based on the Time Bloom Filters, we propose a distributed architecture to enhance the scalability, as shown in Fig. 2.



**Fig. 2.** Distributed architecture for global duplicates filtering

The filtering process works as follows. First, all data pass through the “Distributing Node”, which routes data to appropriate “Filtering Node”. Then the “Filtering Node” determines whether this data is redundant. The routing process is shown in Algorithm 2 in Java language. First we use a hash function to compute a node id for the RFID data. Here we use the multiplication method, as shown in line 2. Then we send the data to corresponding “Filtering Node” identified by the id. The time complexity of Algorithm 2 is  $O(1)$ . In the server side, we can't use Algorithm 1 to filter duplicates because of the network latency. Here, we use the Time Bloom Filter proposed by Lee and Chung [8] to filter the duplicates, as shown in Algorithm 3 in Java language. The filtering process in Algorithm 3 is very similar to Algorithm 1. We first compute  $k$  hash values for the tag id in line 1 to line 2. Then check whether this data is received  $\tau$  time before in line 3 to line 7. Finally we set the corresponding positions to the timestamp of this new data in line 8 to 9. The time complexity of Algorithm 3 is  $O(k)$ .

Algorithm 2. *Distributing\_Node(RFID\_Data x)*  
Input:  $x$  is triple (ReaderID, TagID, Timestamp)  
1:  $tag\_id = x.TagID;$   
2:  $node\_id = \lfloor d * (tag.id * A \bmod 1) \rfloor;$   
3: send  $x$  to the Filtering Node indentified by  $node\_id$ .

Algorithm 3. *Filtering\_Node(RFID\_Data x)*  
Input:  $x$  is triple (ReaderID, TagID, Timestamp)  
Output: true or false (whether  $x$  is a duplicate)  
1:  $for(i = 0; i < k; i++)$   
2:      $p_i = hash_i(x.TagID);$   
3:     boolean duplicate = true;  
4:      $for(i = 0; i < k; i++)$   
5:          $if(BL[p_i] = 0 \text{ or } x.Timestamp - BL[p_i] > \tau)$  then  
6:              $duplicate = false;$   
7:             break;  
8:      $for(i = 0; i < k; i++)$   
9:          $BL[p_i] = x.Timestamp;$   
10:     return duplicate;

**Theorem 1.** The false positive error rate of the Distributed Time Bloom Filters is  $(1 - (1 - \frac{1}{m})^{kn/d})^k$ , where  $d$  is the number of filtering nodes,  $k$  is the number of hash functions,  $m$  is the number of positions in each Time Bloom Filter, and  $n$  is the number of RFID data record sent to the server within  $\tau$  time unit.

**Proof.** Suppose the hash function in Algorithm 2 can map the data into the filtering node uniformly, then there're about  $\frac{n}{d}$  data sent to each filtering node within  $\tau$  time unit. According to the proof of the false positive error rate of Time Bloom Filter [8], the false positive error rate in each filtering node is  $(1 - (1 - \frac{1}{m})^{kn/d})^k$ . Since the filtering nodes in our architecture work independently, the total number of miss-filtered elements is  $\frac{n}{d} \cdot (1 - (1 - \frac{1}{m})^{kn/d})^k \cdot d$ , which is just  $n \cdot (1 - (1 - \frac{1}{m})^{kn/d})^k$ . So the overall false positive error rate is  $(1 - (1 - \frac{1}{m})^{kn/d})^k$ .

## 4 Experimental Results

In this section, we use the detection model in [11] to generate synthetic RFID data sets for experiments. Four readers are used to monitor each location. The local duplicate filtering module is simulated by software implementation. The distributed Time Bloom Filters in the server are implemented using multi-threading in Java. We conducted our experiments on a Dell PowerEdge 11G server running windows 2003.

We consider three key characteristic of our method: the bandwidth used to transfer the RFID data, the processing rate, and the false positive error rate.

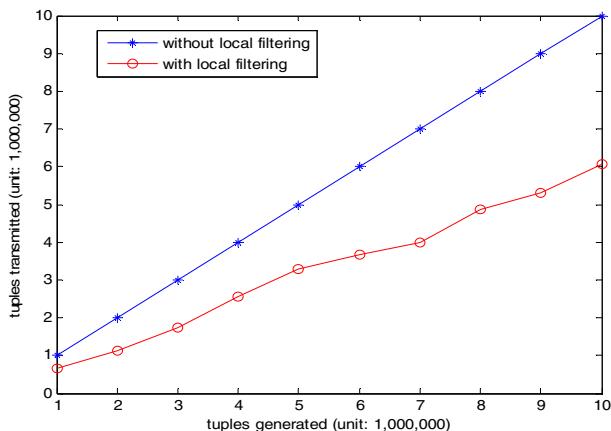
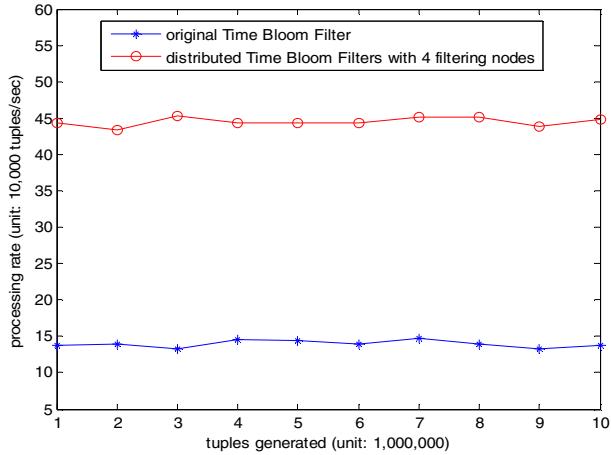
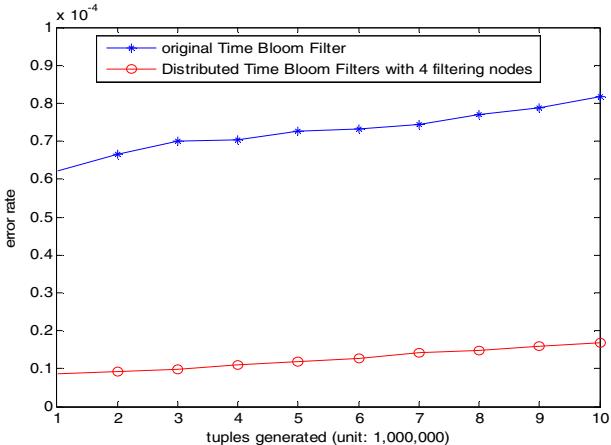


Fig. 3. Bandwidth reduced with local filtering

The number of tuples transferred from readers to the server is used to measure the consumed bandwidth. Fig. 3 shows the number of tuples transferred according to the total number of generated tuples. Without local filtering, all the tuples will be transmitted over the network to the server. This is corresponding to the line with cross in Fig. 3. By adding a local duplicate filtering module, all the local duplicates will be filtered in the readers, thus saving lots of network bandwidth. The height difference between the two lines in Fig. 3 represents the bandwidth reduced. In this synthetic data set, about thirty percent of the data is local duplicates.

**Fig. 4.** Processing rate comparison**Fig. 5.** Error rate comparison

The processing rate is measured by the number of tuples processed per second. The original Time Bloom Filters introduced in [8] are compared with our distributed Time Bloom Filters, which consists of one distributing node and four filtering nodes. The result is shown in Fig. 4. When using four filtering nodes, our method is about three times faster. There's some overhead in the distributing node and the communication between distributing node and the filtering node. Finally, we consider the false positive error rate of our approach. Here we compare our method with the Time Bloom Filter [8]. The result is shown in Fig.5. The error rate of our method is much lower than the original Time Bloom Filters as we expected. This is because we used three more computing nodes than the original Time Bloom Filter. However, adding a few more computing nodes in the server side is easy and won't cost much.

## 5 Conclusions and Future Work

In this paper, we propose a two-layer approach to filter duplicates in RFID data streams. Experimental results show that our method works well. When the network bandwidth is scarce resource and the local duplicate rate is high in applications, our method works best. In the future, we plan to integrate the open source distributed computing project Hadoop into RFID duplicate filtering in the server side.

**Acknowledgement.** The authors would like to thank the anonymous reviewers for their valuable comments and suggestions. This work is supported in part by the National Natural Science Foundation of China under Grant No. 61170035, Jiangsu 973 project (No. BK2011022), National Natural Science Foundation of Jiangsu under Grant No. BK2011702, the Special Grants of China Postdoctoral Science Foundation (No. 200902517), the special project of Nanjing Scientific Committee Foundation (No. 020142010), the Qing Lan Project Foundation of Jiangsu Province 2010, the Star of Excellence Zijin Foundation of NJUST 2009.

## References

1. Martinez-Sala, A.S., Egea-Lopez, E., Garcia-Sanchez, F., Garcia-Haro, J.: Tracking of Returnable Packaging and Transport Units with Active RFID in the Grocery Supply Chain. *Computers in Industry* 60(3), 161–171 (2009)
2. Harrop, P.: RFID in the Postal Service. More RFID (2005)
3. Derakhshan, R., Orlowska, M., Li, X.: RFID Data Management: Challenges and Opportunities. In: IEEE International Conference on RFID, pp. 175–182 (2007)
4. Bloom, B.H.: Space/Time Tradeoffs in Hash Coding with Allowable Errors. *Communications of ACM* 13(7), 422–426 (1970)
5. Metwally, A., Agrawal, D., Abbadi, A.E.: Duplicate Detection in Click Streams. In: 14th International Conference on World Wide Web, pp. 12–21 (2005)
6. Deng, F., Rafiei, D.: Approximately Detecting Duplicates for Streaming Data Using Stable Bloom Filters. In: Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data, pp. 25–36 (2006)
7. Wang, X., Zhang, Q., Jia, Y.: Efficiently Filtering Duplicates over Distributed Data Streams. In: International Conference on Computer Science and Software Engineering (CSSE), pp. 631–634 (2008)
8. Lee, C.H., Chung, C.W.: An Approximate Duplicate Elimination in RFID Data Streams. *Data & Knowledge Engineering* 70(12), 1070–1087 (2011)
9. Bai, Y., Wang, F., Liu, P.: Efficiently Filtering RFID Data Streams. In: VLDB Workshop on Clean Databases (2006)
10. Broder, A., Mitzenmacher, M.: Network Applications of Bloom Filters: a Survey. *Internet Mathematics*, 485–509 (2002)
11. Jeffery, S.R., Garofalakis, M., Franklin, M.J.: Adaptive Cleaning for RFID Data Streams. In: 32th International Conference on Very Large Data Bases, VLDB, pp. 163–174 (2006)

# RFID Uncertain Data Cleaning Framework Based on Selection Mechanism

Xiufeng Xia, Lijuan Xuan, Xiaoming Li, and Ying Li

School of Computer Shenyang Aerospace University Shenyang, China  
{xiaxiufeng, xuanlijuan1115}@163.com,  
jackmtlee@yahoo.cn, qd365@foxmail.com

**Abstract.** Radio frequency identification (RFID) is one of the key technologies in Internet of Things. The mass and uncertainty of RFID raw data limit the development of the technology seriously. Through the analysis of the uncertain data, an RFID data cleaning framework based on selection mechanism (CFBS) is established. The framework introduces selection mechanism, and can select the optimal cleaning lines according to the cleaning nodes' judgment conditions. It reduces the delay generated by data transmission and cleaning, because it needn't travel all the cleaning nodes in cleaning framework. The experimental results show that the cleaning framework can ease the pressure of data transmission, and improve the efficiency of data cleaning greatly.

**Keywords:** RFID, selection mechanism, uncertain data, cleaning framework.

## 1 Introduction

The Radio Frequency Identification (RFID) is an automatic identification technology, originated in 1990s. And a typical RFID system consists of three parts: tag, reader and antenna. First of all, the readers send RF signal to their workspaces through the antenna, the tags in readers' workspace are activated after receiving the signal; then, tags return response signal to readers through antenna; finally, readers explain the response signal, which aims at identifying and capturing the tags information.

Though, the RFID system has been widely used in logistics assembly, manufacture, traffic management, etc. However, the data management issues appear at the same time. Compared with the traditional data forms, RFID data has its own characteristics, such as simplification, time-space relativity, redundancy and semantic richness [1]. In addition, RFID data has another two important characteristics, one is mass, and the other is uncertainty [2] which leads to a phenomenon that the raw data accuracy captured by readers is only 60-70% [3-4]. Unfortunately, the data quality is so poor that the data can't be applied for upper application systems directly.

According to the characteristics of RFID data, an RFID data cleaning framework based on selection mechanism is established. Through considering the data mass and uncertainty, it is found that the framework can clean the data streams selectively and ease the congestion caused by mass data, and then the efficiency of cleaning is enhanced. Finally, the experiments indicate that the framework not only reduces the cleaning time cost, but also improves the cleaning real-time effectively.

## 2 Related Work and Analysis

In RFID system, the data becomes uncertain because of environment and RF signal physical characteristic, etc. According to the characteristics, the uncertain data is divided into positive reading, negative reading and redundant reading.

- 1) Positive reading: The tags, which aren't present at a reader's workspace, but do be read by the reader.
- 2) Negative reading: The tags, which appear in reader's workspace truly, but are not read by the reader at all.
- 3) Redundant reading: Redundant reading is divided into redundant reader and redundant data. The former means that a tag is captured by two readers at least at the same time; the latter refers to the large number of duplicate records.

Usually, the positive reading and redundant reader are random and have a smaller number. Both of them are affected by the environment easily. For the negative reading, it is a common phenomenon in RFID data and holds the largest proportion in RFID uncertain data.

At present, most of experts and scholars focus on the design of cleaning algorithms[5-7]but ignoring uncertain data characteristics, and the cleaning framework of RFID data is involved much less. In [8], a scalable sensor data processing system ESP is proposed. The system is based on sliding window and introduces the time and space, and uses query language to clean the data step by step. Although the framework is also applied to the RFID system, still can't avoid the issue that sliding window size is difficult to determine.

The characteristics of RFID uncertain data are so various that it is impossible to clean data with a unified strategy. Therefore, most of researchers view the layered cleaning as the research foundation. Through the analysis of data characteristics, a layered cleaning framework based on selection mechanism (CFBS) is established in this paper, and the framework may give the optimal cleaning path for RFID data according to the judgment conditions.

## 3 Cleaning Framework Based on Selection Mechanism

For RFID system, the data gives a great support for upper applications. But the RFID raw data accuracy is only 60-70%, so the data must be cleaned in order to output the higher quality data.

### 3.1 Problem Description

For RFID raw data, the data format is a simple structure:  $\langle EPC, Reader, Timestamp \rangle$ , it means that the reader captures the EPC (Electronic Product Code) at timestamp. To describe the cleaning process better, several concepts will be defined as follows.

Definition 1: (Interaction). Interaction is a communication process between tag and reader. It says that the reader sends RF signal to the workspace, and then the tag

returns a response after receiving the RF signal. In this paper it is denoted as  $IC_{i,j}^k$  ( $k = 1, 2, \dots, n, n \in N^*$ ).

And  $i$  denotes the tag numbers,  $j$  denotes the reader numbers,  $IC_{i,j}^k$  refers to the  $k$ TH time interaction completed by  $R_j$  and  $Tag_i$ .

**Definition 2:** (Processing Unit). Processing unit is defined as a set of finite number of interactions, and it is also the smallest cleaning unit. The relation between processing unit and interaction is expressed as formula (1).

$$PU_i^m = \{IC_{i,j}^1, IC_{i,j}^2, IC_{i,j}^3, \dots, IC_{i,j}^k, \dots, IC_{i,j}^{count}\} \quad (1)$$

In formula (1),  $i$  denotes the tag numbers;  $j$  denotes the reader numbers; and  $k$  denotes the  $k$ TH time interaction between  $R_j$  and  $Tag_i$ ;  $count$  denotes the tag response times in processing unit;  $m$  denotes the  $m$ TH processing unit of  $Tag_i$ ; and  $k = 1, 2, 3, \dots, count, count \in N^*$ . If the tag is dynamic, the reader numbers may be variable.

In order to clean data in the form of processing units, this paper will do a pretreatment on RFID raw data. The method is: the raw data will be classified according to the tag numbers when gained from the central database; then the data will be grouped in a certain size and the groups are called processing units which have their own numbers. After the simple pretreatment, the attributes  $Time\_fir$ ,  $Time\_last$  and  $Count$  are extracted, and the data storage form is as (2) below.

$$< EPC, Reader, Time\_fir, Time\_last, Count > \quad (2)$$

And  $Time\_fir$  denotes the timestamp that the tag responds reader first time;  $Time\_last$  denotes the timestamp that the tag responds reader last time;  $Count$  denotes the response times.

**Definition 3:** (Cleaning Node). Cleaning node is the link of cleaning RFID data. In this paper, the cleaning nodes are divided into positive reading cleaning node, redundant reading cleaning node and negative reading cleaning node.

**Definition 4:** (Cleaning Queue). Cleaning queue is the orderly arrangement of processing units which is waiting for being cleaned in cleaning node. Correspondingly, cleaning queue also divided into positive reading cleaning queue, redundant reading cleaning queue and negative reading cleaning queue, denote as  $P\_CQ$ ,  $R\_CQ$ ,  $N\_CQ$  respectively.

Take  $P\_CQ$  as an example, the processing units  $PU_i^m, PU_i^{m+1}, PU_j^k, PU_j^{k+1}$  all contain positive reading, so these processing units will enter into  $P\_CQ$  to wait for being cleaned. The  $P\_CQ$  is as (3) below.

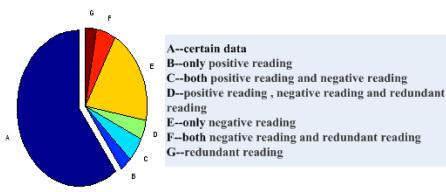
$$P\_CQ = \{ \dots, PU_i^m, PU_i^{m+1}, PU_j^k, PU_j^{k+1}, \dots \} \quad (3)$$

In formula (3),  $i, j$  denote the tag numbers;  $m, k$  denote the processing unit numbers.

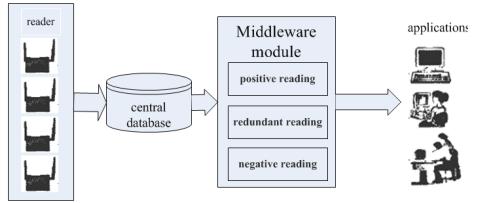
### 3.2 Cleaning Framework Based on Selection Mechanism

Mass is one of RFID data characteristics. However, since the RFID systems demand the real-time higher and higher, we must minimize data cleaning time in order to transfer the data into the upper application with the fastest speed.

For RFID data, the little part of the data contain positive reading, negative reading and redundant reading at the same time, and most of them only contain one or two kinds of uncertain data, RFID data distribution is shown as Fig.1.



**Fig. 1.** Distribution of RFID data

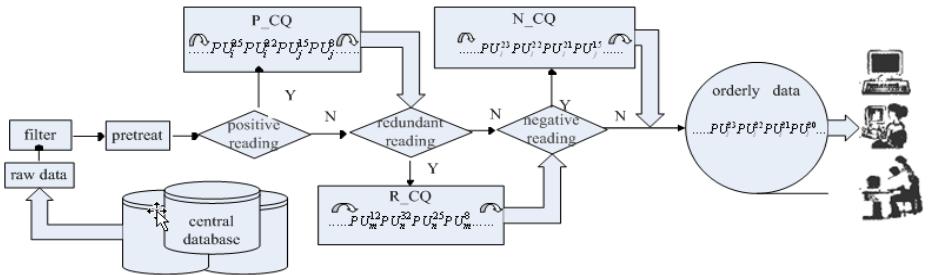


**Fig. 2.** The traditional cleaning framework

As shown in Fig.1, negative reading is the most common phenomenon in uncertain data, positive reading and redundant reading are lesser and random.

The traditional method inputs all the data into the cleaning framework directly without any detecting, that is, all of raw data will travel every link of cleaning framework. Therefore, it brings a huge pressure to data transmission, and the traditional cleaning framework is shown as Fig.2.

Although the traditional strategy can ensure the integrity of the cleaning to some extent, the characteristics of data are ignored. When the RFID data is mass, the efficiency of cleaning is affected seriously since the latter data is turn to be cleaned only when the front data cleaning is completed. According to the shortages above, an RFID data cleaning framework based on selection mechanism is proposed, and shown as Fig.3.



**Fig. 3.** RFID data cleaning framework based on selection mechanism

In CFBS, the data enters into the framework in the form of processing units after pretreatment. When the processing unit arrives at a new cleaning node, the system will judge whether it should be added into the cleaning queue according to the

cleaning conditions, if true, the processing unit will enter the cleaning queue; else, it will be outputted to the next cleaning node to judge. Through the selection mechanism, the processing units will be shunted to the most appropriate cleaning node. And in this way the processing units are not necessary to pass all the cleaning nodes, and the cleaning time cost is reduced effectively.

### 3.3 Processing Flow of CFBS

- Step1: Gain the raw data from the central database and transmit them to the filtering layer to filter out the dirty data;
- Step2: Receive the data from Step1 and do the pretreatment. First ,class them according to tag numbers; then , group the same kind of tags in a certain size for processing units; finally, transmit the processing units to the positive readings cleaning node;
- Step3: Receive the data from Step2 and detect whether the processing unit contains positive reading. If false, go to Step5;
- Step4: The processing unit will enter the positive reading cleaning queue to wait for being cleaned;
- Step5: Receive two aspects of data: 1) the raw data without positive reading; 2) the data outputted from the positive reading cleaning queue. Detect whether the processing units contains redundant reading (in this paper, it means redundant reader).if not true, that means

$$\neg \exists Tag (Tag \in R_i \& Tag \in R_j \text{ } (i \neq j) \text{ at the same time})$$

then go to Step7;

- Step6: The processing units will enter the redundant reading cleaning queue to wait for being cleaned;
- Step7: Receive all processing units entered into the negative reading cleaning node, and detect whether the processing units contains negative reading. If false, go to Step9;
- Step8: The processing units will enter the negative reading cleaning queue to wait for being cleaned;
- Step9: Receive all processing units and sort them in order to ensure the data ordered in the global;
- Step10: Submit the data to the upper applications.

Through the cleaning process, the RFID data have became smooth and continuous. In this paper, different cleaning strategies are applied in different cleaning nodes, we will no longer give unnecessary details because of the limit of the length.

The delay generated by data transmission and cleaning is the bottleneck of the efficiency in RFID systems, so when the cleaning framework is designed, we must minimize the data transmission time cost. In this paper, the cleaning framework based on selection mechanism not only ensures the effective of cleaning, but also decreases the time cost.

## 4 Experimental Results and Analysis

In the section, we will test and verify the effectiveness of RFID uncertain data cleaning framework based on selection mechanism. The experimental environment is the Intel Pentium Dual E2180 2.00GHz processor, 2G memory, operating the Windows XP2 Professional system platform, database MySql5.0. Execute the CFBS with Java language.

### 4.1 Experimental Data

In this paper, simulated datasets are chosen as the experimental data source. First, we select a random function  $\text{Random}(\text{float time}, \text{int value})$ , and the random function generates the random numbers (value) with a fixed frequency (time), which is used to simulate the readers to send signals to their workspaces regularly; then, we sort the random numbers and delete the repeat records; finally, in order to make simulated data similar to the real data maximum, we will adjust the uncertain data proportion of simulated datasets. The experimental data are divided into different datasets, the datasets size are 1000, 5000, 10000, 20000. The experimental data is shown as table 1.

**Table 1.** The experimental data

Data name	Data size
DataSet1	1000
DataSet2	5000
DataSet3	10000
DataSet4	20000

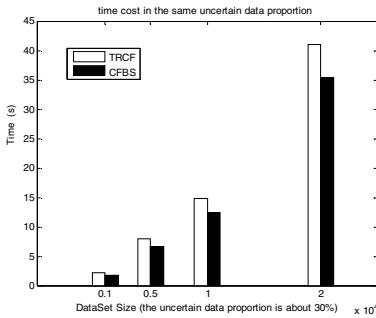
### 4.2 Experimental Result

In order to compare the time performance between traditional cleaning framework (TRCF) and CFBS comprehensively, this paper will measure the cleaning time cost between them.

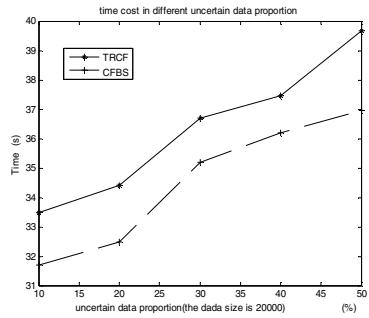
Firstly, when the data size is different but the uncertain data proportion is the same, we set the uncertain data proportion to be about 30% in datasets. The experimental result is shown in Fig.4.

As shown in Fig.4, when dataset size increases, the time cost of TRCF and CFBS increases, too. Therefore, the data cleaning time has a linear relationship with dataset size roughly. In addition, with the increasing of data size continuously, the gap of time cost will also increase between TRCF and CFBS.

Secondly, when the data size is the same but the uncertain data proportion is different, we set the data size is 20000, the time cost is shown in Fig.5.



**Fig. 4.** Time performance comparison in the same uncertain data proportion between TRCF and CFBS



**Fig. 5.** Time performance comparison in different uncertain data proportion between TRCF and CFBS

As shown in Fig.5, when the data size is the same, the time cost will increase with the increasing of uncertain data proportion. In other words, the cleaning time cost is affected by not only data size but also uncertain data proportion.

Through the two experiments above, it is found that the CFBS proposed in this paper takes an advantage in time performance. For CFBS is based on selection mechanism, it needn't travel all cleaning nodes like traditional framework but selects the optimal cleaning path according to the cleaning conditions. In this way, time of data transmission reduces; the time of waiting for being cleaned reduces, too. And with the increasing of dataset size, CFBS advantage will become more obvious.

## 5 Conclusion

This paper proposed an RFID data cleaning framework based on selection mechanism (CFBS), which can improve the efficiency of RFID data cleaning greatly. According to the characteristics of uncertain data, the framework can solve the delay problem generated by data transmission and cleaning. Through the analysis of the experiments, it is shown that the framework has a good expansibility and can reduce the time cost and delay of data cleaning. In order to get the high-quality data better, we will pay more attention to the optimization of cleaning strategies in the further research.

## References

1. Gu, Y., Yu, G., Zhang, T.: RFID Complex Event Processing Techniques. *Frontiers of Computer Science and Technology* 1(3), 255–267 (2007)
2. Xu, J., Yu, G., Gu, Y., Wang, Y.: Uncertain Data Management Technologies in RFID. *Frontiers of Computer Science and Technology* 3(6), 561–576 (2009)
3. Fishkin, K.P., Jiang, B., Philipose, M., Roy, S.: I Sense a Disturbance in the Force: Unobtrusive Detection of Interactions with RFID-tagged Objects. In: Davies, N., Mynatt, E.D., Siio, I., et al. (eds.) *UbiComp 2004. LNCS*, vol. 3205, pp. 268–282. Springer, Heidelberg (2004)

4. HaHnel, D., Burgard, W., Fox, D., et al.: Mapping and Localization with RFID Technology. In: International Conference on Robotics and Automation, pp. 1015–1020 (2004)
5. Bai, Y., Wang, F., Liu, P.: Efficiently Filtering RFID Data Streams. In: The First International VLDB Workshop on Clean Databases (CleanDB) Workshop, Seoul, Korea, pp. 50–57 (2006)
6. Jeffery, S.R., Garofalakis, M., Franklin, M.: Adaptive Cleaning for RFID Data Streams. In: Proceedings of the 32nd International Conference Very Large Data Bases (VLDB), Seoul, Korea, pp. 163–174 (2006)
7. Carbunar, B., Ramanathan, M.K., Koyutuk, M., Hoffmann, C., Grama, A.: Redundant Reader Elimination in RFID System, Sensor and Ad Hoc Communications and Networks. In: IEEE SECON 2005, 2005 Second Annual IEEE Communications Society Conference, Santa Clara, California, USA, pp. 176–184 (September 2005)
8. Jeffery, S.R., Alonso, G., Franklin, M.J., et al.: A Pipelined Framework for Online Cleaning of Sensor Data Streams. In: Proceedings of the 22nd International Conference on Data Engineering (ICDE), Atlanta, Georgia, USA, pp. 773–778 (2006)

# An Effective Temporary ID Based Query Tree Algorithm for RFID Tag Anti-collision

Yun Tian<sup>1</sup>, Gongliang Chen<sup>1</sup>, and Jianhua Li<sup>1,2</sup>

<sup>1</sup> School of Information Security Engineering,  
Shanghai Jiaotong University, Shanghai, China

<sup>2</sup> Department of Electronic Engineering,  
Shanghai Jiaotong University, Shanghai, China  
`{ruth_tian,chengl,lijh888}@sjtu.edu.cn`

**Abstract.** This paper proposes an effective temporary ID based query tree algorithm (TID QTA) for RFID tag anti-collision. Tags with 96-bit UID are mainly concerned. In TID QTA, a 16-bit sequence is selected from tag ID as its TID and query tree algorithm is implemented based on TID. If two tags have the same TID, they will renew their TIDs by selecting another 16 bits from their IDs. The simulation results show that the average number of transmitted bits for one tag identification in TID QTA is the fewest among QTA and its variants.

**Keywords:** anti-collision, tree-based algorithm, temporary ID, tag identification.

## 1 Introduction

Radio Frequency Identification (RFID) systems have been widely used in industries, retailing, daily life, etc. RFID systems consist of tags and readers where readers identify tags via wireless communication. When several tags respond to a reader's interrogation at the same time, a collision occurs. Because of the collision, the reader fails to identify tags and has to inquire repeatedly. In order to improve identification efficiency, anti-collision protocols are required to arbitrate when collisions occur.

There are three types of tag anti-collision algorithms in RFID systems: ALOHA based algorithms, tree based algorithms and hybrid algorithms[1]. In ALOHA based algorithms, tags transmit their IDs at a randomly selected time in order to avoid responding to the reader simultaneously[2]. In tree based algorithms, collided tags are splitted into two groups iteratively until no collision occurs and reader can identify one tag at a time[3-6]. ALOHA based algorithms are simpler because tags do not need to do the prefix matching while tree based algorithms offer full identification of tags because they won't cause the tag starvation problem[7]. Hybrid algorithms combine the advantages of ALOHA based and tree based algorithms together[1].

In a query tree algorithm (QTA)[3], the reader sends a binary string to tags. If the prefix of a tag ID matches this inquiring string, the tag will transmit the rest of its ID to the reader. When only one tag ID matches the prefix, the reader can identify this tag. If there is a collision, the reader adds 0 and 1 at the end of the previous inquiring

string and identifies the tags with these two prefixes of IDs respectively. QTA is memoryless because tag doesn't need to record the position of the inquired bit. The only computation required for each tag is to match its ID against the binary string in the query. However, QTA suffers if tag IDs are long, for example, 96 bits in EPC. The inquiring string is initialized to be one bit and is lengthened during the interrogation. In the worst case, the inquiring string would be as long as 95 bits. This leads to long delay in the identification. Thus new approaches should be adopted in order to control identification delay when tag IDs are long.

RN16QTA[4] shortens the transmitted bits by using a 16-bit random number as tag's temporary ID to implement query tree algorithm. Inspired by RN16QTA, this paper proposes a novel temporary ID based query tree algorithm (TID QTA) in order to decrease the identification delay when the size of tag ID is large. Different with RN16QTA, our algorithm selects 16 bits from tag ID as its TID and begins a round of QTA interrogation according to TIDs. If two tags have the same TID, they will renew their TIDs by selecting another 16 bits from their IDs. Then a new round of QTA will be implemented. The simulation results show that TID QTA outperforms QTA and its variants because the average number of transmitted bits for one tag identification in TID QTA is the fewest and thus improve the identification efficiency.

The rest of this paper is organized as follows: in section 2, we briefly review several tree based anti-collision algorithms. Then a novel temporary ID based query tree algorithm, TID QTA, is proposed in section 3. In section 4, performance evaluation of TID QTA and other tree based algorithms is presented. Finally, conclusion is drawn in section 5.

## 2 Related Work

In this section, several variants of QTA will be introduced.

A collision tree protocol (CT)[5] uses Manchester code in QTA to detect the first collided bit in the tags' responses. Then the next query is lengthened directly to the first collided position. Thus, CT decreases the query times and improves the efficiency of tag identification. RN16QTA applies the characteristics of EPC Class 1 Gen. 2 protocol, i.e. the generation of a 16-bit pseudo random number (RN16), to QTA. RN16 works as a temporary ID (TID) of the tag in QTA. Since the size of tag ID is 96 bits in EPC, both the inquiring and responding bits will be reduced by using a 16-bit TID. But RN16QTA would fail to identify all the tags when some tags hold the same TID. ERN16QTA[6] solves this problem. In ERN16QTA, the 96-bit EPC ID is splitted into 6 16-bit blocks, namely  $EPC_1, EPC_2, EPC_3, EPC_4, EPC_5, EPC_6$ . Let  $TID_0 = RN16$ . When two or more tags have the same  $TID_i$ ,  $TID_{i+1}$  is generated where  $TID_{i+1} = TID_i \oplus EPC_{i+1}$ . Here,  $\oplus$  denotes bit-wise XOR operation. A new round of QTA will start to inquire the tags with  $TID_{i+1}$ . Since  $EPC_i = TID_i \oplus TID_{i-1}$ , the identified tag only needs to transmit the rest of its ID and the reader can recover the whole ID from the previous communication. Thus ERN16QTA not only solves the problem of undistinguishable RN16, but also decreases the transmitted bits.

However, we think that the XOR operation is redundant and may lead to failure. The recovery of  $EPC_1, \dots, EPC_i$  needs  $TID_0, \dots, TID_i$ . The same result can be obtained if  $EPC_i = TID_i$ , which means that the tag can just use  $EPC_i$  as  $TID_i$ . In

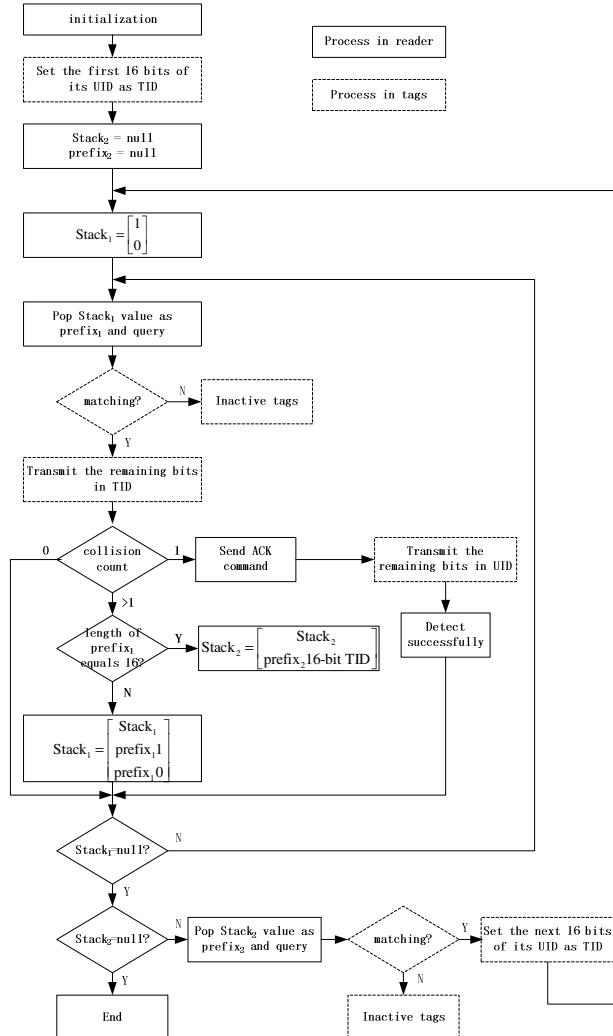
addition, assume that tag A holds  $\text{RN16}_1$  as its TID and tag B  $\text{RN16}_2$ . If  $\text{RN16}_1$  and  $\text{RN16}_2$  are both undistinguishable, and  $\text{RN16}_1 \neq \text{RN16}_2$ . Then tag A and tag B will be both inquired during the query of  $\text{TID}_1$ . It may happen that  $\text{RN16}_1 \oplus \text{EPC}_{\text{A}1} = \text{RN16}_2 \oplus \text{EPC}_{\text{B}1}$ , where  $\text{EPC}_{\text{A}1}$  and  $\text{EPC}_{\text{B}1}$  refer to the  $\text{EPC}_1$  of tag A and tag B respectively. Thus, tag A and tag B won't be identified during the query of  $\text{TID}_1$  either. If the rest bits of tag A and tag B are the same, the reader will fail to identify tag A and tag B until the end of the algorithm. For example, let  $\text{RN16}_1 = (0F5A)_{16}$ ,  $\text{RN16}_2 = (A5AA)_{16}$ . When  $\text{EPC}_{\text{A}1} = (C396)_{16}$  and  $\text{EPC}_{\text{B}1} = (6966)_{16}$ , then  $\text{TID}_{\text{A}1} = \text{RN16}_1 \oplus \text{EPC}_{\text{A}1} = \text{RN16}_2 \oplus \text{EPC}_{\text{B}1} = \text{TID}_{\text{B}1}$ , which means that in the query of  $\text{TID}_1$ , the reader will fail to identify tag A and tag B. If the rest bits in tag A and tag B are the same,  $\text{TID}_{\text{A}i}$  will always be the same as  $\text{TID}_{\text{B}i}$ . The result is that the reader cannot identify these two tags.

In this paper, we present an effective TID based QTA by selecting 16 bits from tag UIDs as TIDs. Our algorithm doesn't require tags to generate random numbers. Moreover, tag with the unique TID will only transmit the remaining bits in its UID rather than the whole 96 bits.

### 3 TID QTA

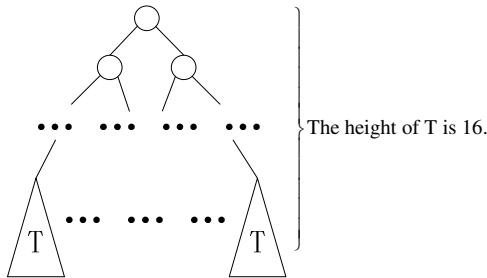
Fig. 1 shows the flow chart of our TID QTA. During the initialization, each tag will set a 16-bit sequence as its TID. TID is usually selected as the most different part of tag IDs. For example, as described in [6], a common situation may be the identification of large amount of products manufactured by the same company. The first 60 bits of products' 96-bit UID are the same: encoding scheme (header: 8 bits), company prefixes (GMN: 28 bits) and object class (24 bits). The last 36 bits in UID stand for the serial number which indicates the unique product. Thus, the proper choice of TID is the last 16-bit sequence in UID. But it is possible that two or more tags have the same TID, and this will lead to the failure of identification of these tags. We solve this problem by initializing a stack, denoted as  $\text{Stack}_2$  in the reader during the initialization phase.  $\text{Stack}_2$  is used to store the undistinguishable TIDs. At the beginning of TID based QTA, it is null.  $\text{Prefix}_2$  is used as the request which is the top value from  $\text{Stack}_2$ . It is also initialized to be null. Then the QTA begins. The reader sends a binary string as  $\text{prefix}_1$ . If its TID matches  $\text{prefix}_1$ , the tag will transmit the rest of its TID back to the reader. If the length of  $\text{prefix}_1$  is equal to 16 bits, which means that there is no bit in the rest of tag's TID, then the tag will transmit an arbitrary binary bit. When only one tag responds, the reader can get the TID of this tag. It then sends a 2-bit ACK together with the 16-bit TID. So the tag which has the same 16-bit TID transmits the rest of its UID to the reader. Therefore, this tag is identified. If a collision occurs and  $\text{prefix}_1$  is shorter than 16 bits, the reader will push  $\text{prefix}_1 + "0"$  and  $\text{prefix}_1 + "1"$  into  $\text{Stack}_1$  for the further request. When  $\text{prefix}_1$  is already 16-bit in length, which means that there are at least two tags that have the same TID, the reader pushes  $\text{prefix}_2 + \text{prefix}_1$  into  $\text{Stack}_2$ . QTA will be iterated until  $\text{Stack}_1$  becomes null. When  $\text{Stack}_1$  is null, which means that this round of QTA is finished,  $\text{Stack}_2$  will be checked. If  $\text{Stack}_2$  is not null, the reader will pop the value as  $\text{prefix}_2$ . Since the tags whose UIDs match  $\text{prefix}_2$  are undistinguishable during the last QTA interrogation, they will renew their TIDs by the next 16-bit sequence in UIDs and try to be identified in the next

QTA. Other tags change the state into inactive state, i.e. they won't respond during the next round of QTA interrogation. If Stack2 is null, that means all the tags have been identified and the identification process is finished.



**Fig. 1.** The flow chart of TID QTA

Fig. 2 illustrates the identification tree in TID QTA. The height of the tree is 16, which is the length of TID. Each identification tree refers to a round of QTA interrogation. The tree is constructed in a recursive way. Each leaf in the tree is a T which presents one of the following three types: null, a leaf node indicating a tag, or a tree like the identification tree. In the current QTA, tags with the same TID are in the same T, which means that these tags cannot be identified in the current round of QTA. A new round of QTA will be done according to that T.

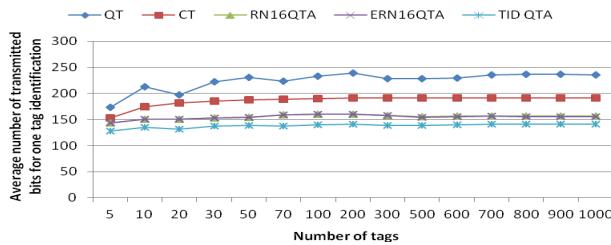


**Fig. 2.** The identification tree in TID QTA

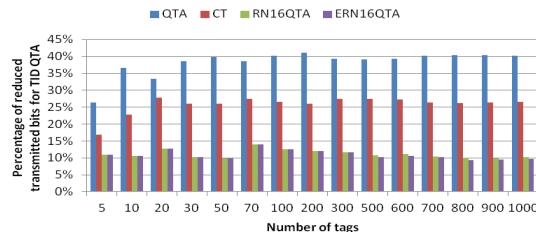
## 4 Performance Evaluation

The simulation is intended for the performance of TID-based QTA compared with that of QTA and other variants including CT, RN16QTA and ERN16QTA.

The simulation program is written with Java and our simulating scenario assumes that there is only one reader in the interrogation field. The number of tags increases from 5 to 1000. The length of tag IDs is 96 bits. They are distributed uniformly. In RN16QTA, when two or more tags generate the same RN16, new RN16 will be generated.



(a) Average number of transmitted bits for one tag identification



(b) Percentage of reduced transmitted bits for TID QTA

**Fig. 3.** Performance Evaluation

Fig.3 (a) shows the average number of transmitted bits for one tag identification. As shown in the figure, our TID-based QTA needs fewest bits to identify one tag.

This is because that tags in TID QTA responds to reader's request with the remaining bits of 16-bit TID rather than the remaining bits of whole 96-bit UID in QTA and CT. So tag's responded bits are reduced in TID QTA by using shortened TID as UID. Moreover, each bit transmitted in TID QTA carries information of tag UID. But in RN16QTA and ERN16QTA, the query and response of RN16 have nothing to do with tag UID. So, after a tag has been identified in RN16QTA, the tag has to transmit its 96-bit UID to the reader, while in TID QTA, the part of tag UID served as TID has already been transmitted to the reader and the tag only has to send the remaining bits of UID. Thus, the average number of transmitted bits for one tag identification in TID QTA is fewer. Fig.3 (b) shows the percentage of reduced transmitted bits for TID QTA compared with QTA and other variants. The average reduction percentages compared with QTA, CT, RN16QTA and ERN16QTA are 38.23%, 25.82%, 11.15% and 10.96% respectively.

## 5 Conclusion

In this paper, we present a temporary ID based query tree tag anti-collision algorithm called TID QTA. Unlike RN16QTA, the TID in our algorithm is chosen from the tag UID rather than generating a new random number. When several tags are undistinguishable because they have the same TID, new TIDs will be chosen. The simulation results show that TID QTA is more efficient than QTA and its variants including CT, RN16QTA and ERN16QTA because the transmitted bits during the identification process are reduced.

**Acknowledgments.** This work is supported by the National Natural Science Foundation of China under Grant No. 61071078 and the National Basic Research Program of China under Grant No. 2010CB731403.

## References

1. Klair, D.K., Chin, K.-W., Raad, R.: A Survey and Tutorial of RFID Anti-collision Protocols. *IEEE Communications Surveys & Tutorials* 12(3), 400–421 (2010)
2. EPC™ Radio-Frequency Identity Protocols Class-1 Generation-2 UHF RFID Protocol for Communications at 860MHz - 960MHz version 1.2.0,  
[http://www.gs1.org/gsmp/kc/epcglobal/uhfc1g2/uhfc1g2\\_1\\_2\\_0-standard-20080511.pdf](http://www.gs1.org/gsmp/kc/epcglobal/uhfc1g2/uhfc1g2_1_2_0-standard-20080511.pdf)
3. Law, C., Lee, K., Siu, K.-Y.: Efficient Memoryless Protocol for Tag Identification (Extended Abstract). In: 4th International Workshop on Discrete Algorithm and Methods for Mobile Computing and Communications, pp. 75–84. ACM, New York (2000)
4. Choi, J.H., Lee, D., Lee, H.: Query Tree-based Reservation for Efficient RFID Tag Anti-collision. *IEEE Communications Letters* 11(1), 85–87 (2007)
5. Jia, X., Feng, Q., Ma, C.: An Efficient Anti-collision Protocol for RFID Tag Identification. *IEEE Communications Letters* 14(11), 1014–1016 (2010)
6. Yang, C.-N., He, J.-Y.: An Effective 16-bit Random Number Aided Query Tree Algorithm for RFID Tag Anti-collision. *IEEE Communications Letters* 15(5), 539–541 (2011)
7. Zhu, L., Yum, T.-S.P.: A Critical Survey and Analysis of RFID Anti-collision Mechanisms. *IEEE Communications Magazine* 59(5), 214–221 (2011)

# An Anti-collision Algorithm of RFID Tags Based on CDMA

Weijun Zhang<sup>1</sup>, Shuping Zhang<sup>1</sup>, and Dawei Zhang<sup>2</sup>

<sup>1</sup> School of Computer, Shenyang Aerospace University,  
110000 Shenyang, China

<sup>2</sup> School of Information Science and Technology, Eastern Liaoning University Dandong,  
118000 Dandong, China  
xping0211 @163.com

**Abstract.** Radio frequency identification (RFID) is rapidly come into our lives in recent years, and it is widely used in the areas of logistics, production, and transportation and so on. The technology of anti-collision is a key of RFID. Most of the anti-collision algorithms proposed before are based on TDMA. When the number of tags is very large, it can not well solve the problem. In this paper, an anti-collision algorithm based CDMA is proposed, using the m sequence as the spreading code and combined with the S-ALOHA algorithm that is known as CMS-ALOHA algorithm. This paper analyzes the throughput of CMS-ALOHA, and gives the mathematical analysis. Experimental results show that this algorithm has good performance.

**Keywords:** RFID, anti-collision algorithm, CMS-ALOHA, CDMA.

## 1 Introduction

Radio frequency identification is a non-contact automatic identification technology, with the characteristic of high accuracy, far reading distance, large amount of data storage and strong durability and so on, which is widely used in the area of logistics, production, and transportation and so on. The tag is the information carrier of RFID [1]. When multiple tags appears in the reader's reading range simultaneously, information sent by each tags will be mixed together, and thus cause conflict, resulting in the data wrong or lose even missed read. Therefore, it is necessary to design an effective anti-collision algorithm to solve this problem.

The current main anti-collision algorithms are ALOHA algorithm and binary tree search algorithm [2]. They are all based on Time Division Multiple Access (TDMA). In some applications, such as logistics, since the number of RFID tags is relatively large, its cost is required to be quite low, and the memory and computing are also very limited, so it can not perform complex calculations. Some of the criterions limit the communication bandwidth of RFID system. Therefore, the number of information bits transmitted between the reader and tags should be diminished. Because of these limitations, this paper proposes to use the technology of Code Division Multiple Access (CDMA) to solve conflicts.

## 2 The CDMA Technology

The CDMA (Code Division Multiple Access) allows all users use all the bands simultaneously, and regarded the signals that send by other users as noise. We do not have to take into account the problem of signal collision. In this technique, the stream of information to be transmitted is divided into small pieces, each of which is allocated across to a frequency channel across the spectrum. The technology of CDMA identifies different potential users by different code sequences using the orthogonal or quasi-orthogonal of code sequences.

The spreading sequence codes in RFID system is usually generated by the m sequence, Gold sequence or Walsh sequence. Its autocorrelation function and cross-correlation function of selected sequences must satisfy the following formulas [3]:

$$R_m(\tau) = \frac{1}{L} \sum_{j=1}^L m(i, j)m(i, j + \tau) = \begin{cases} 1 & \tau = 0 \\ 0 & \tau \neq 0 \end{cases} \quad (1)$$

$$\rho_m(\tau) = \frac{1}{L} \sum_{j=1}^L m(i, j)m(k, j + \tau) = 0 \quad (2)$$

Where  $L$  is the length of spreading code,  $i \in \{1, N\}$ ,  $N$  is the number of objects to be identified.

In all of the pseudo-random sequences, m sequences is with good pseudo-random characteristic and usually used as the spread code in the spread communication system. It can constitute the orthogonal code set of the orthogonal codes spread spectrum system. Therefore, this article selected m sequence as spreading sequence. The generation of m sequence [4]: as the state with all “0” is unable to transfer to other states, through the appropriate tap feedback and modulo 2 adders, n-class linear shift register can produce the maximum possible period of sequence as  $p = 2^n - 1$ .

The autocorrelation function of m sequence is:

$$R(\tau) = \begin{cases} 1 & \tau = 0 \\ -1/p & \tau = 1, 2, \dots, p-1 \end{cases} \quad (3)$$

Where  $p$  is the cycle of code sequences. The autocorrelation function of m sequence peaked at  $\tau = 0$ .

The cross-correlation of m sequence is the similarity measure of two different code sequences. When the code sequence is used to distinguish addresses, the selected code must be the sequence with wee value of cross-correlation function to avoid mutual interference among users. In the CDMA system, the value of cross-correlation function should be as small as possible, which is convenient to distinguish different users, or has strong anti-interference. The cross-correlation function of m sequences is with the property of multi-value. The two m sequences which meet the following conditions have better cross-correlation function feature [4]:

$$R_{xy}(\tau) \leq \begin{cases} 2^{(n+1)/2} & n \% 2 \neq 0 \\ 2^{(n+2)/2} & n \% 2 = 0 \& n \% 4 \neq 0 \end{cases} \quad (4)$$

### 3 The Principle of CMS-ALOHA Algorithm

#### 3.1 The Slotted-ALOHA Algorithm

Slotted-ALOHA algorithm is referred to as S-ALOHA algorithm [5], which is improved base on the ALOHA algorithm. The basic principle of ALOHA algorithm is that each user can send data at any time if only need. It can also be applied in RFID systems. As soon as the tags come into the reader's range, they will send their message to the reader. The reader detects the received signals and judges whether the received signal conflicts with others. If so, the reader sends the command which stops the tag to send information. The tag waits for a while randomly before sending next time. If not, the reader begins to communicate with the tag.

The average amount of the exchange of data packets in  $T$  time is:

$$G = \sum_{n=1}^n \frac{\tau_n}{T} r_n \quad (5)$$

Average throughput:

$$S = G e^{-2G} \quad (6)$$

When  $G = 0.5$ , the maximum value of  $S$  is 18.4%.

S-ALOHA algorithm [6] is the improvement of ALOHA algorithm, which divided the time into time-slots  $\tau$  equal to a packet length, whose length is fixed. But it ruled that starting sending time must be the starting point of the time-slot, which can halve the conflict, because only the messages send in the same slot may conflict. However, when it is needed to resend, S-ALOHA retransmission algorithm must be delayed the time equal to integer multiple of time-slot.

The average throughput of this algorithm is:

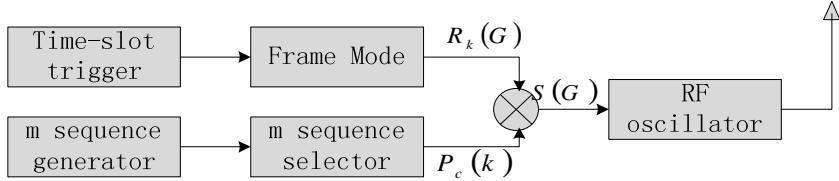
$$S = G e^{-G} \quad (7)$$

When  $G = 1$ , the maximum value of  $S$  is 36.8%, which is the 2 times of the ALOHA algorithm. When  $G > 1$ , the system is unstable, and they do not meet the needs if there are a large number of tags in the reader's area.

#### 3.2 The Principle of CMS-ALOHA Algorithm

The main idea of CMS-ALOHA algorithm is summarized as follows: there are two sets of data stored for each tag, which are its own ID number and a random m sequence of spreading code that written in the production. The actually signal tags send is the spread spectrum signal. The reader get the real tag ID number through disspreading, and the reader can prevent a part of collisions using the orthogonal of spreading codes to improve the throughput of the system. However, the number of tags ID may be much greater than the number of spreading codes. So there will be a case that different tags select the same spreading code in practice.

In addition, all of the spreading code groups are pre-selected and limited the quantity, and stored in the reader. The block diagram of CMS-ALOHA system is as the fig.1 shows:



**Fig. 1.** The block diagram of CMS-ALOHA system

The tags information is shown as  $R_k(G)$ , and the m sequence selector is shown as  $P_c(k)$ , then gets the spread spectrum signal  $S(G)$  is:

$$S(G) = R_k(G)P_c(K) \quad (8)$$

Then, the spread spectrum signal  $S(G)$  modulated by the carrier is sent to the reader. The reader gets the information that tags send by corresponding demodulation.

## 4 The Mathematical Analysis of CMS-ALOHA Algorithm

Assuming CMS-ALOHA system generate the m sequence by n-level shift registers, it can produce maximum of different states of as  $p = 2^n - 1$ , that is n-linear feedback shift registers may have the longest period equal to  $2^n - 1$ , the spreading sequence code is expressed as  $0, 1, \dots, p-1$ . Suppose the tag  $i$  use the spread-spectrum sequence  $l$  in a time-slot ( $l$  is a random integer,  $0 \leq l \leq p-1$ ) to send its own ID number, but if there are other tags using the same spreading code to send its ID number in this time-slot, it will cause conflict, and the information can not be received by reader correctly. In the time slot, the probability of a RFID tag using a spreading code to send information is:

$$P(l) = 1/p, 0 \leq l \leq p-1 \quad (9)$$

If there are  $K+1$  tags sending their own ID numbers in the  $T$  time slot, and the probability that the other  $K$  tags do not use the same spreading codes with the tag  $i$  is:

$$P(K) = \sum_{l=0}^{p-1} P(l)[1 - P(l)]^K = 1 - \frac{1}{p} \quad (10)$$

The process of the initial information access of the information arrival can be viewed as a Poisson process, and in a time slot, the probability of  $K$  tags arrived is:

$$R_K = G^K e^{-\frac{G}{K!}} \quad (11)$$

Here, the load  $G$  is the average arrival number in a Reader's identification area between  $T$  times. In the beginning of each time slot, assuming  $K+1$  access information in a Reader's region wish to send their information, and  $N$  access information can be received by the reader at least. Let  $P_c(K)$  is the probability of a tag's information successfully captured and recovery under  $K$  disturbances, and the average channel throughput  $S$  is the number of average information sent successfully and received successfully by Readers within each time slot, it can be denoted as:

$$S(G) = \sum_{K=0}^{\infty} \min(K+1, N) R_{K+1}(G) P_c(K) \quad (12)$$

In the CMS-ALOHA system, if multiple access interference and Gaussian white noise are not considered, then  $P_c(K) = P(K)$ , at the same time  $S(G)$  can also be expressed as:

$$\begin{aligned} S(G) &= e^{-G} \sum_{K=0}^{\infty} \frac{G^{K+1}}{(K+1)!} \left[ 1 - \frac{1}{N} \right]^K \min(K+1, N) \\ &= e^{-G} \sum_{K=0}^{\infty} \frac{G^{K+1}}{(K+1)!} \left[ 1 - \frac{1}{N} \right]^K (K+!) = G e^{-G} \sum_{K=0}^{\infty} \frac{\left[ \left( 1 - \frac{1}{N} \right) G \right]^K}{K!} = G e^{-\frac{G}{N}} \end{aligned} \quad (13)$$

When  $G = N$ ,  $\max(S) = N/e = G/e$ . It can be seen clearly that the throughput of CMS-ALOHA algorithm is proportional to the order of spreading code, and is  $N$  times for the maximum throughput (0.36) of S-ALOHA algorithm. Due to the impact of hardware, the order of spreading codes is limited, and can not improve the system's throughput by increasing an infinite number of the order of spreading code.

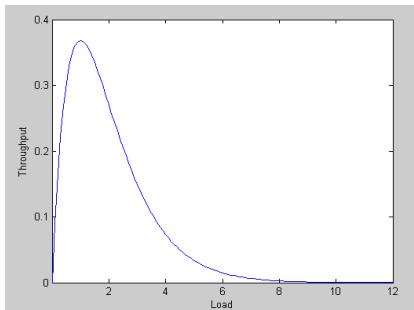
## 5 The Simulation of Computer

In order to validate the CMS-ALOHA algorithm, we simulated the performance with the MATLAB. The simulation condition is the tags subject to Poisson distribution. Fig.2 shows the throughput of S-ALOHA algorithm, that is the throughput of CMS-ALOHA algorithm when  $N = 1$ . It can be seen from the figure that the simulation result is consistent with the theoretical result, and the maximum throughput is 0.36.

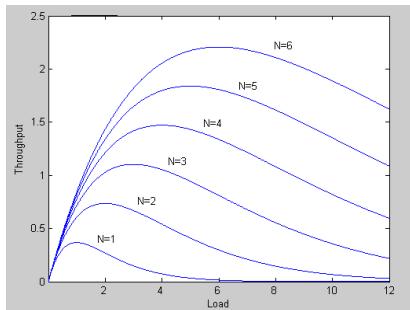
Compare with the throughput of different order spreading codes, shown in Fig.3, the experiment demonstrates that: when  $N = G$ , the tag reaches the maximum throughput; when  $N \neq G$ , the throughput of tags are less than the maximum.

In order to achieve our goals, we cannot increase the  $N$  unrestricted. That's because of in the actual production; it will be limited by the hardware circuit. The experimental results verify the CMS-ALOHA algorithm is  $N$  times than the maximum throughput (0.36) of S-ALOHA algorithm.

The results show that the throughput of CMS-ALOHA algorithm is better than S-ALOHA algorithm. From the two figures, we see that the experimental result is consistent with the theoretical analysis.



**Fig. 2.** The throughput of S-ALOHA algorithm



**Fig. 3.** The comparison of the throughput for different order of spreading codes

## 6 Conclusions

This paper proposes the CMS-ALOHA anti-collision algorithm with  $m$  sequence as spreading sequences. It combines the CDMA with the S-ALOHA algorithm, which can effectively solve the problem of RFID tags confliction when there are a large number of tags in a Reader's area. Theoretical and experimental results show that, the throughput of CMS-ALOHA algorithm is better than S-ALOHA algorithm. And the more of the  $m$  sequences' order, the greater the throughput of the system will be. Compared with the popular S-ALOHA algorithm, CMS-ALOHA algorithm greatly increases the throughput of the RFID system.

However, for the value of  $N$ , this paper did not discuss in-depth, because this subject should be considered to specific environmental constraints, we can make a detailed analysis in the specific applications.

## References

1. Mutti, C., Floerkemeier, C.: CDMA-based RFID Systems in Dense Scenarios: Concepts and Challenges. In: 2008 IEEE International Conference on RFID, The Venetian, Las Vegas, Nevada, USA, April 16-17 (2008)
2. Shin, W.J.: A Capture-Aware Access Control Method for Enhanced RFID Anti-collision Performance. IEEE Communications Letters 13(5) (May 2009)
3. Yu, S., Peng, Y., Zhang, J.: RFID Anti-collision Algorithm merging Security Mechanism. Advanced Materials Research 216 (2011)
4. Tang, Z., He, Y.: Research of multi-access and anti-collision protocols in RFID systems. In: IEEE International Workshop on Anticounterfeiting, Security, Identification, Xiamen, China, pp. 377–380 (2007)
5. Zhen, B., Kobayashi, M., Shimizu, M.: Framed ALOHA for multiple RFID objects identification. IEICE Trans. Commun. E88-B(3) (March 2005)
6. Mazurek, G.: RFID System With Spread-Spectrum Transmission. IEEE Transactions on Automation Science and Engineering 6 (2009)

# The Research on Electronic Tag Quantity Estimate Arithmetic Based on Probability Statistics

Lin Zhou<sup>1</sup>, Zhen Li<sup>2</sup>, Yingmei Chen<sup>1</sup>, and Tong Li<sup>3</sup>

<sup>1</sup> Xi'an Communications Institute, Xi'an 710106, China

Zhou8201@163.com

<sup>2</sup> Army Unit 68036, Xi'an 710061, China

<sup>3</sup> Xi'an University of Architecture and Technology, Xi'an 710055, China

**Abstract.** In the RFID system, the precise estimation of electronic tag number is very important for increasing tag identification precision and efficiency. The paper introduces three estimate arithmetic based on probability statistics, and analyzes and compare the arithmetic performance through mathematic analyze and simulation experiment. The result shows that the ZE estimate arithmetic, which estimate based on zero frame, and the CE estimate arithmetic, which based on conflict frame, can precisely estimate the number of tags under certain system load, and it also indicates that the max tag number which CE can estimate is bigger than ZE with the same system time slots, and the difference is a linear relationship with the system time slots.

**Keywords:** RFID, electronic tag, probability statistics, number estimate, algorithm research.

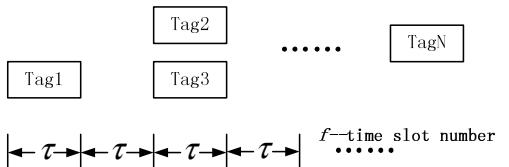
One of the important tasks of the RFID (Radio Frequency Identification) system is to identify a large number of electronic tags. Nowadays, most researches are focused on the tag identification algorithm, the estimation method of the number of tags are really less. However, the inaccurate estimation of the total number of tags is the main reason for inefficient recognition process of tags [1]. To solve this problem, the paper [1] proposed an estimation algorithm based on probability and statistics, and it qualitatively analyzed the characteristic of the algorithm. Our work is also based on the probability and statistics, but additionally we analyzed the effective and the maximal estimation tags of the algorithm quantitatively, finally we provide some simulations to support our opinions. Our work provides a reference for the selection of tag estimation algorithm. In this paper below we firstly introduce the system model, describe and analyze the algorithm, and then we analyze the maximal estimation tags of algorithm and simulate the algorithm's result.

## 1 System Model

It is often need to identify a large number of tags in RFID system, and the identification algorithms can usually be divided into two categories, which is ALOHA

algorithm and tree algorithm [2], however, these two algorithms have one thing in common: they all need an accurate estimation of the number of electronic tags, which used as a parameter for the identification algorithm to shorten the identification time. In ALOHA algorithm, only when the time slot number equal to the tag quantity, then the system reach the maximum throughput [3]. For tree algorithm, the number of tags can guide the algorithm to set the number of time slot of the recognition stage [4]. Therefore, there should be a tag number estimation phase before tag identification, and the whole identification process is composed of two parts: the estimation stage and the recognition stag.

Assume that we use passive tags [5-7] in RFID system, and the tags send feedback information to device in fixed time slot ALOHA model [8] at the tag quantity estimation stag. It means that when the tag receives the signal from the identification device, it send the feedback signal in a time slot which randomly selected from the  $f$  slots. After that, the identification device detect each time slot, if it is only selected by a label, we call it single tag slot; If there is no label, then empty time slot; If there are two labels at least, we call it conflict slot, shown in Fig 1. The tag number estimation which based on probability and statistics is realized through statistics on various types of slot number, which is the foundation of estimation.



**Fig. 1.** Electronic tags random select the time slot

The label's feedback signature does not need to be unique, and it's only have to make the recognition device can monitor conflict in a time slot at the estimation stage. The time for estimation stage will be far less than the recognition stage [1]. Therefore, the quantity estimation for tags can either improve the efficiency of identification or ensure the timeliness of the system.

## 2 Algorithm Description

We label the system's time slot number with  $f$ , and the tag randomly selects a time slot which we label as  $j(1 \leq j \leq f)$  to send data. So when the identification device detecting the slot, there will be 3 situations:

- (1) None of the tags select the time slot, which called empty slot. We use a random variable “ $X_j$ ” to express the situation, if it is empty then  $X_j=1$ , otherwise  $X_j=0$ .
- (2) There is only one tag selected the time slot, which called single tag slot. And we use a random variable “ $Y_j$ ” to express the situation, and  $Y_j=1$  represent the situation,  $Y_j=0$  for others.

(3) There are two tag selected the time slot at least, which called the conflict time slot. And we also use a random variable “ $V_j$ ” to denote the case, when  $V_j=1$  it means the case take place, and  $V_j=0$  for other cases.

The equation  $X_j + Y_j + V_j = 1$  is true for every time slot “ $j$ ” for only one of the situation will take place.

In order to facilitate the analysis, we use random variable  $N_0$ ,  $N_1$  and  $N_c$  to represent the total number of empty slots, single slots and conflict slots in the  $f$  time slots, respectively. Therefore we have:

$$N_0 = \sum_{j=1}^f X_j \quad (1)$$

$$N_1 = \sum_{j=1}^f Y_j \quad (2)$$

$$N_c = f - N_0 - N_1 \quad (3)$$

Assume that the system load is defined as the ratio of the tag number and slot number, that is  $\rho = t/f$ , then the expected function value of  $N_0$ ,  $N_1$  and  $N_c$  can be approximately expressed as [1]:

$$E[N_0] \approx fe^{-\rho} \quad (4)$$

$$E[N_1] \approx f\rho e^{-\rho} \quad (5)$$

$$E[N_c] \approx f(1 - (1 + \rho)e^{-\rho}) \quad (6)$$

This is the expected value of different slot type, and we can get the estimation value of tag number  $\hat{t}$  by solving the equation based on the statistics of the slots and the system time slots which is known. If we use the random variables  $n_0$ ,  $n_1$  and  $n_c$  to represent the number of empty slot, single slot and conflict slot respectively and there exist:

$$E[N_0] \approx fe^{-(t_0/f)} = n_0 \quad (7)$$

$$E[N_1] \approx f(t_1/f)e^{-(t_1/f)} = n_1 \quad (8)$$

$$E[N_c] \approx f(1 - (1 + t_c/f)e^{-(t_c/f)}) = n_c \quad (9)$$

By solving the above three equations, we can get three estimation value of the tag number, they are  $\hat{t}_0$ ,  $\hat{t}_1$  and  $\hat{t}_c$ . Thus, by the statistics of the slot number of the different situations, we can get three estimation methods, estimation of empty time

slots, referred to as ZE; estimation of single-label slot, referred to as SE; and conflict slot estimation method, referred to the CE.

The three estimation algorithm based on probability and statistics have different characteristics. From the preceding analysis, we known the normalized expected value  $E[N_0]/f$ ,  $E[N_1]/f$  and  $E[N_c]/f$  are the function about the system load, and their relationships shown in Fig2.

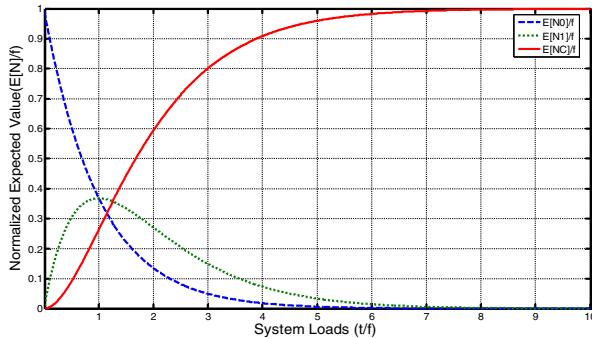


Fig. 2. The normalized expected value to different system loads

From the figure above we can see that when the system loads is low, there are many empty time slots, but the number of single-label slot and multi-label slot is very small; when the load increases, the number of empty time slots reduce and the single-label slot and conflict slot increase. The number of single-label slot reach the maximum point at the system loads  $\rho = 1$ , and since it then begin to decrease. It can also be found that the number of empty time slot and multi-label slot are monotonic to the system loads, so there will always be only one estimation number of tags to the given system loads and statistic tag number. According to the statistics of a single time slot, there are two different system loads corresponding to it except the spot at  $\rho = 1$ . Therefore, the ultimate label estimation values are not unique, and it can not be used to the estimation of label number. Thus, we analyze the estimation method based on the empty time slots and multi-label slot in the follow.

### 3 The Maximum Estimate Number of the Algorithm

When the number of tags is much larger than the number of time slots ( $t \gg f$ ), the probability of time slot with multiple labels will be very high, and in that case, the estimation of empty label slot and multi-label slot will be invalid for the corresponding slot quantity is very small and some times be zero, so we can not get a definite estimation of the tag numbers. Therefore, it is equal to say that the ZE and CE algorithm are both applicable under appropriate system loads. We can see that from Fig 1 the maximum system loads for algorithm CE is about 6, and 4 for the algorithm ZE. When the total number of tags is fixed, and gradually increase the system time

slots, reducing the system loads, the exact estimate probability of the tags will increase. That is, for a specific estimation algorithm, a given number of time slots  $f$  and the probability  $\theta$  ( $\theta < 1$ ), we can always find a maximum number of tags, which can be exactly estimated using the algorithm with the probability bigger than  $\theta$ .

We adopt probability and statistics method to find the maximum number of tags the algorithm can estimate. If we assume the maximum number of tags to be  $t$ , the system time slot fixed to be  $f$ , for the ZE algorithm the probability of no empty time slot to be less than  $(1 - \theta)$ ,  $\Pr[N_0 = 0] < (1 - \theta)$ . Under the same assumption above, for CE algorithm the probability for the situation of  $n_c=f$  is less than  $(1 - \theta)$ ,  $\Pr[N_0 = 0, N_1 = 0] < (1 - \theta)$ .

Meanwhile, due to the number of empty time slots and a single-label slots are follow the Poisson distribution [9], and intensity are  $\lambda_0 = fe^{-\rho}$ ,  $\lambda_1 = fpe^{-\rho}$ , respectively, then we have:

$$\Pr[N_0 = 0] \approx e^{-\lambda_0} \leq (1 - \theta) \quad (10)$$

$$\Pr[N_0 = 0, N_1 = 0] \approx e^{-\lambda_0 - \lambda_1} \leq (1 - \theta) \quad (11)$$

And because:

$$\lambda_0 = fe^{-\rho} = fe^{-(t_0/f)} \quad (12)$$

Therefore,  $e^{-\lambda_0}$  is an increasing function about  $t_0$ , so solving the equation:

$$e^{-\lambda_0} = e^{-fe^{-(t_0/f)}} = (1 - \theta) \quad (13)$$

We can get the maximum of  $t_0$ , called  $t_{0\_max}$ , it is the upper bound of the ZE algorithm.

Similarly, for  $\Pr[N_0 = 0, N_1 = 0] \approx e^{-\lambda_0 - \lambda_1}$  due to:

$$\lambda_0 = fe^{-\rho} = fe^{-(t_c/f)} \quad (14)$$

$$\lambda_1 = fpe^{-\rho} = fpe^{-(t_c/f)} \quad (15)$$

The function  $\Pr[N_0 = 0, N_1 = 0] \approx e^{-\lambda_0 - \lambda_1}$  is an increasing function about  $t_c$ , so solving the equation:

$$e^{-\lambda_0 - \lambda_1} = e^{-fe^{-(t_c/f)} - fpe^{-(t_c/f)}} = 1 - \theta \quad (16)$$

We will get the maximum of  $t_c$ , called  $t_{c\_max}$ , it is the upper bound of the CE algorithm.

Joint equation (13) and (16), we will get the relationship of  $t_{c\_max}$  and  $t_{0\_max}$  as follows:

$$t_{c\_max} - t_{0\_max} = f \times \ln(1 + (t_{c\_max}/f)) \quad (17)$$

Because the Maximum system loads which algorithm ZE applicable is definite, so we can find from the equitation that the difference between  $t_{c\_max}$  and  $t_{0\_max}$  are linear relationship about the system loads.

## 4 The Simulation

In front, we have analyzed three tag estimation algorithm based on probability and statistics, found the maximum number of the tags the algorithms can estimate, and we also found their relationships. Now, we will use the simulation method to verify the above results. To determine the validity and accuracy of the algorithm, we set two parameters: the success probability  $\theta$  ( $0 < \theta < 1$ ) and fault tolerance value  $\beta$  ( $\beta > 0$ ), which require the algorithm to get estimation value  $\hat{t}$  which satisfy the equitation  $\hat{t} \in (t - 0.5 * \beta * t, t + 0.5 * \beta * t)$  with the probability greater than  $\theta$  [1], and  $t$  represent the real tag quantity.

### 4.1 Algorithm Validation

To verify the effectiveness of the algorithm, we take the method as follow: fixed the system time slots, gradually increasing the number of tags, and then compare the algorithm's estimation value with the real number of tags. As fig 3 shown the ZE and CE algorithm's estimation results when we set system slots  $f=100$  and gradually increasing the number of tags.

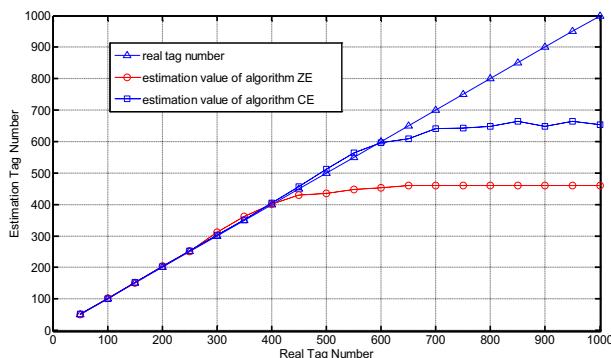


Fig. 3. The relationship between algorithm performance and tag quantity

The figure indicate that when we set the system time slot  $f=100$ , the tag quantity less than 400, the algorithm ZE's estimation value are perfect match to the real tag quantity, the algorithm ZE is effective; For the CE, as long as the number of tags to be less than 600, the estimated value is consistent with the true value. Therefore, you can find the two algorithms are able to accurately estimate the number of tags in a certain system loads, and algorithm CE can accommodate a greater system load than algorithm ZE.

## 4.2 Verify Algorithm's Maximum Estimation Number

Through the above analysis and derivation we know that, given the system time slots, we can calculation the algorithm's theoretical maximum estimation value of tags by using formula (3) and (4). Whether this calculation method is reliable or not need to testify. We can find the algorithm's maximum estimation value for different system time slots through experiment, and then we compare the experiment value and theoretical value to analyze the method's feasibility of the calculation of maximum number of tags. We set  $\beta = 0.1$  and  $\theta = 0.9$ , the result is shown in Fig 4.

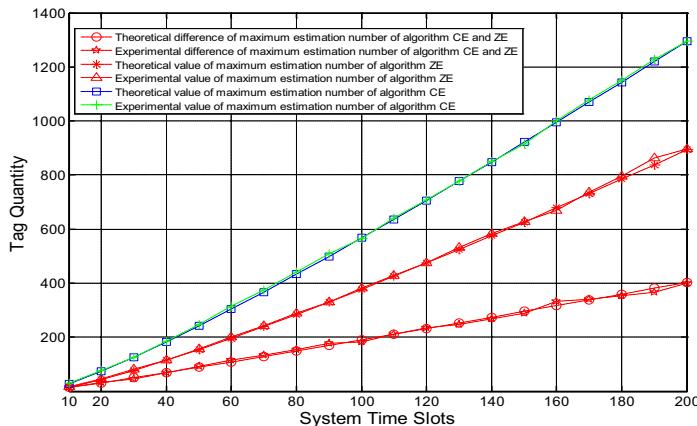


Fig. 4. The relationship between system time slots and algorithm's maximum estimation value

Fig4 shown the theoretical and experimental value of the maximum estimation of algorithm CE and ZE when the system time slots start at 10 to 200, and it also denote the difference of the two algorithms. We can find that the theoretical value and experimental value are consistent, so the algorithm for calculation maximum estimation number of tags is effective and applicable. At the same time, we get the max system loads which algorithm CE and ZE can applicable under different system time slot, shown in Fig 5.

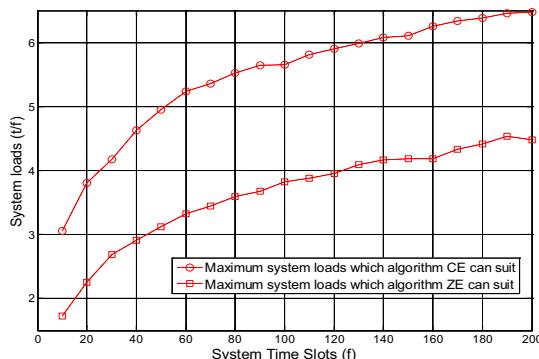


Fig. 5. Max system loads which Algorithm can applicable with different system slots

We can find from Fig 5 that no matter how many system time slots we have, the CE algorithm can adapt bigger system loads compare to the ZE algorithm.

## 5 Summary

To accurately estimate the number of tags has important affection to improve the identification efficiency. For the algorithm based on the single-label slot is not suitable for estimating tag number, article focuses especially on the performance of ZE and CE estimation algorithm. The calculation methods for maximum estimation number of the ZE and CE algorithm are provided through mathematical derivation. The simulation results show that the algorithm ZE and CE can estimate the quantity of tags exactly under a certain system loads, and for the same system slots the maximum estimation number of algorithm CE is bigger than the algorithm ZE, and the difference was a linear relationship with the system slot number.

## References

1. Kodialam, M., Nandagopal, T.: Fast and Reliable Estimation Schemes in RFID Systems. In: MobiCom 2006, Los Angeles, California, USA, pp. 322–333 (2006)
2. Yang, J., Wang, Y., Zhan, Y.: Research on Tag Anti-collision Algorithms of RFID System under Dense Tag Environment. Journal of Sun Yat-sen University 48(6), 147–150 (2009)
3. Schoute, F.C.: Dynamic framed length ALOHA. IEEE Transactions on Communications 31(4), 565–568 (1983)
4. Finkenzeller, K.: RFID Handbook, 3rd edn. John Wiley & Sons, Ltd (2010)
5. Finkenzeller, K.: RFID handbook: Radio frequency identification fundamentals and applications. John Wiley & Sons, Ltd (2000)
6. Want, R.: An introduction to RFID technology. In: IEEE Pervasive Computing, vol. 3, pp. 25–33 (January 2006)
7. Hassan, T., Chatterjee, S.: A taxonomy for RFID. In: Hawaii International Conference on System Sciences, Kauai, HI, pp. 1–10 (January 2006)
8. Tanenbaum, A.S.: Computer networks, 4th edn. Prentice-Hall, New Jersey (2003)
9. Feller, W.: An Introduction to Probability Theory and Its Applications, vol. 1. John Wiley & Sons, Ltd. (1968)

# An Improved RFID Data Cleaning Algorithm Based on Sliding Window

Lingjuan Li<sup>1,2,3</sup>, Tao Liu<sup>1</sup>, Xiang Rong<sup>1</sup>, Jianxin Chen<sup>1</sup>, and Xiaolong Xu<sup>1</sup>

<sup>1</sup> College of Computer, Nanjing University of Posts and Telecommunications,  
Nanjing, Jiangsu 210003, China

<sup>2</sup> Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks,  
Nanjing, Jiangsu, China

<sup>3</sup> Key Lab of Broadband Wireless Communication and Sensor Network

Technology (Nanjing University of Posts and Telecommunications),  
Ministry of Education Jiangsu Province, Nanjing, Jiangsu 210003, China  
{lilj,1010041302,Y004091112,chenjx,xuxl}@njupt.edu.cn

**Abstract.** As one of the important features of RFID systems, the uncertainty lives through the life cycle of RFID application. In order to improve the data cleaning quality for the unstable data of RFID, we analyze the characteristics of RFID data and the possible reason of the uncertainty. Then we improve the SMURF algorithm by adding  $p^*$  and the reading rate to the reading cycle which is coming to the window. In order to test the performance of the improved algorithm, we do several experiments, and the results show that the improved RFID data cleaning algorithm based on sliding window is better than the SMURF when the data stream is unstable.

**Keywords:** RFID, uncertain data, data cleaning, SMURF.

## 1 Introduction

RFID (Radio Frequency Identification) is a kind of non-contact automatic identification technology. It uses two-way radio communication to achieve identification and to exchange data or to read and write record media [1]. In recent years, RFID technology, by right of its features such as flexibility, speediness and wireless identification, gradually attracts people's attention [2], and has been widely used in public safety, production management, mobile tracking, traffic management and other fields.

The data made by RFID is different from that in traditional database and data warehouse. The characteristics of RFID are as follows:

Simple data form of raw data: Data generated from an RFID application can be seen as a stream of RFID tuples of the form (EPC, location, time) [3]. The EPC which can identify an object uniquely is a kind of electronic production code. The Location records the location where RFID reader scans the item, and time is the time when the reading took place [3].

The great amount of data: RFID generates data in an automatic and fast way and often requires detect more than one object. It makes a very large amount of RFID data.

Spatial and temporal nature: RFID applications can observe dynamic changes in the real world, this is its temporal nature; at the same time, RFID can be embedded into the moving objects, this is its spatiality.

Uncertainty: This is an important property of RFID, and the uncertainty has increasingly aroused the attention of scholars. So, we will discuss the types of RFID uncertainty and methods of processing RFID data.

## 2 The Causes of the RFID Data Uncertainty

The factors that cause the RFID data uncertainty are complicated and exist in whole lifecycle of RFID application, and the uncertainty can be divided into objective uncertainty and subjective uncertainty.

In RFID systems, readers communicate with tags by radio waves, so the system is quite vulnerable to environmental impacts, and that may lead to data lost. With the numbers of readers and tags growing, the interference will become particularly serious. That makes the accuracy of the raw data low, typically only 60% to 70%. The uncertainty caused by radio frequency interference is called objective uncertainty, and it mainly includes three aspects: false negative, wrong reading and false positive.

The false negative is lack of reading. It has two meanings. One is that readers cannot read the tags within its detection range; the other is that readers cannot capture the dynamic change of tag entering or leaving the detection range of readers. But that the tag O at time t is not read maybe does not mean false negative [4] because the limitation of the cost and the abnormality of sensing range may make the actual monitoring site not be full covered.

The wrong reading means that the ambient RF noise is treated as tag reaction and is read by tag readers.

The false positive is redundant reading. There are two types of redundancy. One is tag redundancy and the other is reader redundancy. Tag redundancy is the result of reading the tag more than once because of the RF reflect; reader redundancy refers to several readers detect the same tag at the same time because their detection ranges are overlapped.

The simple form and objective uncertainty of RFID data make the RFID data difficult to support advance applications. In recent years, researchers have proposed some data pre-processing technologies for RFID. But the pre-processing may introduce new RFID data uncertainty called subjective uncertainty, which includes: uncertainty of location information, uncertainty of event semantics and uncertainty of event time. Researchers have proposed some algorithms to solve the problem of subjective uncertainty.

This paper will focus on the objective uncertainty of RFID data.

### 3 Cleaning Method Based on Sliding Window

#### 3.1 Cleaning Methods Based on Fixed-Length Sliding Window

Scholars from UCLA proposed a cleaning method of RFID data streams based on fixed-length sliding window [5]. The sliding window is a certain length window and moving with time. Cleaning method based on sliding window is typical and commonly used method. Such method is divided into two types, tuple-based sliding window and time-based sliding window.

The basic principle of the cleaning method is that if there is a tag in sliding window we assume the tag is consisted in the entire window. In this way, the lack of data in the window has been added, and the purpose of data cleaning has done. In addition, the scholars from UCLA proposed the method about wrong reading and redundant reading. They assume that the possibility of wrong reading is much less than the right reading. So in the method a fault threshold is set, and if the number of a tag is less than the threshold, the tag will be marked as a wrong tag. When new tags enter the window, the tag with the amount greater than threshold will be recorded, and then output. The method also provides a redundant reading time: max\_distance. A tag its interval to the same tag is less than max\_distance will be considered as redundant reading, and will not be output.

The above methods only consider the ideal situation of uniform flow, once dealing with the non-uniform data stream, the higher reading rate and larger window size will result in positive reading, and also the lower reading rate and smaller window size will result in negative reading. That can be presented by Fig. 1 [6].

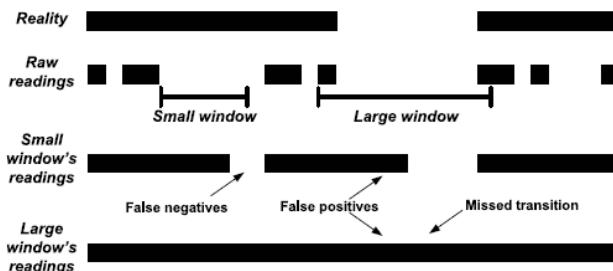


Fig. 1. The problems of sliding window

#### 3.2 The SMURF Algorithm

Taking the inherent non-reliability of RFID data stream into account and aiming at the solving of the problem of determining the window size of the fixed-length sliding window, the scholars from UCB proposed an adaptive RFID data cleaning method – SMURF [6]. The method does not need the parameter of sliding window size; it dynamically adjusts the window size according to average reading rate of the window to increase the accuracy of the data cleaning. If the reading rate of tag is lower, it sets a large window size to reduce negative reading. On the other hand, if the tag reading rate is high, it sets a small window size to reduce positive reading.

The SMURF algorithm proposes that the window size needs to be changed in the situation as follows:

$$|S_i - w_i p_i^{\text{avg}}| > 2 * \sqrt{w_i p_i^{\text{avg}} (1 - p_i^{\text{avg}})}. \quad (1)$$

Where  $S_i$  is times of reading in the window,  $w_i$  is window size;  $p_i^{\text{avg}}$  is the average reading rate of each reading cycle.

When the window size needs to be changed, according to the Bernoulli model, the window size will be set as follows:

$$w_i \geq \lceil \ln(1/\delta) / p_i^{\text{avg}} \rceil. \quad (2)$$

$\delta$  is false negative rate after data cleaning. We can see from the above equation that the window size and the factors that make the window size change are both depended on the core parameter  $p_i^{\text{avg}}$ , and that may cause error in situation of non-uniform data stream.

Let us observe an extreme case, suppose a window contains six reading cycles, the reading rate of the first cycle is 90%, the total reading rate of the second cycle to sixth cycle is 20%, and the reading rate of the new cycle is 90%. When the first cycle is about to be output, the  $p_i^{\text{avg}}$  is not changed, but the reading rate curve has undergone great changes. If we keep the window size unchanged, it will cause a lot of positive reading.

It can be seen that this method can improve the negative and positive reading by changing the unreasonable window size which will cause negative and positive reading, but it cannot completely eliminate them. It also does not take the non-uniform data which sometimes may generate great mistakes into account. So, we make some improvements on SMURF, and propose an improved RFID data cleaning method based on sliding window.

## 4 Design of the Improved Algorithm

### 4.1 Additional Condition of Window Size Change

We can see from the above analysis, when determining whether to change the window size, we should consider not only the average reading rate of window, but also  $p^*$ , which is the reading rate of reading cycle which is about to enter to the window. Therefore, we add a formula as follows for judgment.

$$\omega < \left( \left( \sum_{i=1}^n p_i + p^* \right) / (n+1) \right) - p^*. \quad (3)$$

As an additional condition for original formula (1), the parameter  $\omega$  in this formula is a threshold. When the judgment by original formula failed, this additional condition must take into account. If the right side in the formula is greater than  $\omega$ , the window size should keep to be adjusted.

## 4.2 Determination of the Window Size

As mentioned above, SMURF algorithm adjusts the window size based on the average reading rate of the window. For the poor performance caused by non-uniform data stream, we present a new algorithm, which takes  $p^*$  as one of the factors which can determine the window size. The improved algorithm considers both the average reading rate and the impact of the new reading cycle data coming into the window.

When processing data using sliding window, the tag will be output once the tag appearing in the window. So, the false negative will appear only when the tag is missed in all reading cycle, and there is formula as follows:

$$(1 - p^{avg})^w \leq \delta. \quad (4)$$

$p_i^{avg}$  is average reading rate;  $\delta$  is false negative rate after data cleaning.

The data format of the RFID is <EPC, Location, Time>. Each reading data share the same memory  $C$ . There is  $m$  reading data in reading cycle, so the total cost of memory is  $c * m * p^*$ . The limitation of memory shared by window is  $B$ .

$$c * m * p^* * w \leq B. \quad (5)$$

After integrating formula (4) and formula (5), we get the following inequality:

$$(1 - p^{avg})^w c * m * p^* * w \leq \delta * B. \quad (6)$$

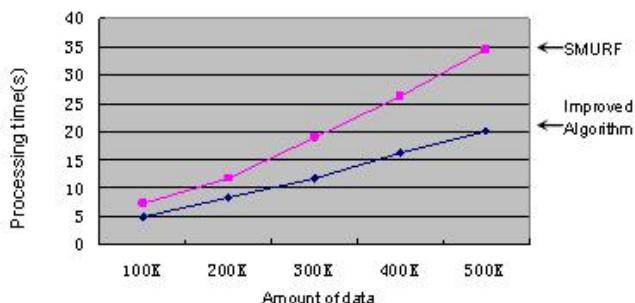
## 5 Performance Analysis of the Improved Algorithm

In order to analyze the performance of the proposed improved algorithm, we compare the improved algorithm with the SMURF algorithm and the cleaning algorithm based on fixed-length sliding window. We compare their cleaning results and time efficiency in normal case (i.e. data does not changing obviously) and the extreme case (i.e. data stream is extremely unstable) by experiments.

The experiments were carried out in Eclipse environment. The experimental procedures can be divided into four parts: RFID data generator, Data reformer, Data cleaning and Result comparator.

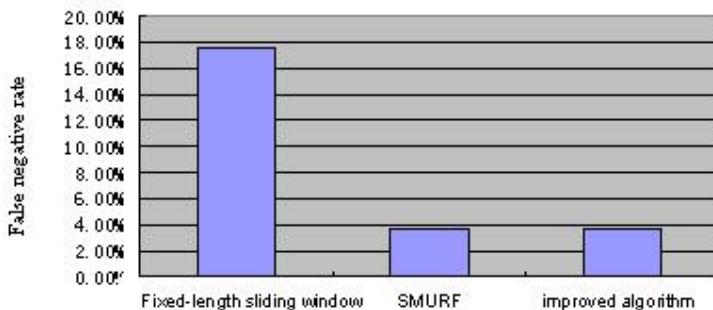
RFID data generator is responsible for producing the RFID real data in the format of <EPC, Location, Time>; Data reformer is responsible for reforming the data generated by data generator in accordance with demand, and produces desired missed reading; Data cleaning is responsible for using various algorithms for data cleaning; Result comparator is responsible for comparing the cleaning results with real data to calculate the false negative rate after cleaning.

The experimental results are shown in Fig. 2 to Fig. 4.



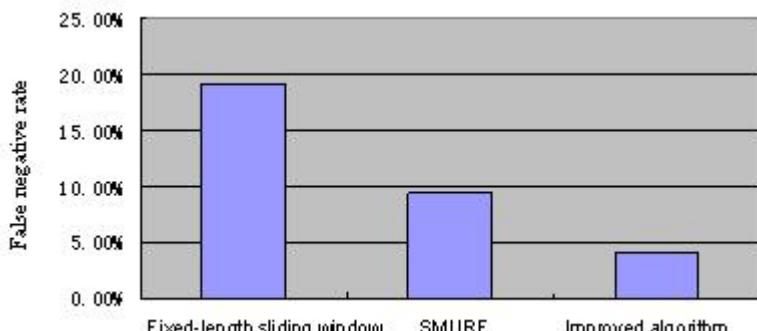
**Fig. 2.** The comparison of time efficiency

Fig. 2 shows the time curves of SMURF algorithm and the improved algorithm when they are dealing with the same data. The result indicates that the improved algorithm can improve both the cleaning efficiency and the time complexity.



**Fig. 3.** The comparison of the cleaning efficiency in normal case

Fig. 3 shows the comparison result of error rates of three cleaning algorithms under normal case (data does not change obviously). It can be seen that SMURF algorithm and improved algorithm all have large improvements compared with fixed-length window, but the improved algorithm has slightly improvement comparing to SMURF.



**Fig. 4.** The comparison of the cleaning efficiency in extreme case

Fig. 4 shows the cleaning efficiency of three algorithms under extreme case (data stream is extremely unstable). Obviously, cleaning efficiency of SMURF is significantly affected, and that of improved algorithm is affected little. This is because the improved algorithm introduces a new parameter  $p^*$ ; in this way it can reduce the impact of non-uniform data stream and improve the effect of data cleaning. In fact the improvement of the improved algorithm for non-uniform data stream is more apparent than that for the normal data stream.

It is obvious that the proposed improved algorithm is good for instable data stream, and can meet the requirements of higher quality cleaning.

## 6 Conclusions

In recent years, as people pay more attention to data, the uncertain data management becomes a hot research issue. RFID as a key wireless technology for development of productive forces, its inherent uncertainty makes new challenges to data management. In this paper, in order to improving the quality of data cleaning, we improve the data cleaning algorithm SMURF; design the additional condition of window size change and the method of determining the window size. By analysis and experimental validation, we can give the conclusion that for stable data stream cleaning the improved algorithm can achieve the same quality with SMURF, and for instability data stream cleaning it can achieve higher quality.

**Acknowledgments.** This paper is supported by the National Basic Research Program of China (973 Program: No.2011CB302903) and the Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (No.yx002001).

## References

1. Jian, S.: Research of Computer Simulation in RFID Test. Shanghai Jiaotong University (2008)
2. Want, R.: The magic of RFID. ACM Queue, 40–48 (2004)
3. Derakhshan, R., Orlowska, M.E., Li, X.: RFID Data Management Challenges and Opportunities. In: IEEE First International Conference on RFID, Texas, pp. 175–182 (2007)
4. Xu, J., Yu, G.: Uncertain Data Management Technologies in RFID. Journal of Frontiers of Computer Science and Technology, 561–577 (2009)
5. Bai, Y., Wang, F., Li, P.: Efficiently filtering RFID Data Streams. In: The First International VLDB Workshop on Clean Databases Workshop, Seoul, Korea, pp. 50–57 (2006)
6. Jeffery, S.R., Garofalakis, M., Franklin, M.: Adaptive Cleaning for RFID DATA Streams. In: Proceedings of the 32nd International Conference on Very Large Data Bases, Seoul, Korea, pp. 163–174 (2004)

# Methods to Recognize Special Tags in UHF RFID System

Lei Hu, Zhen Huang, and Bowen Chen

School of Electronic Information, Wuhan University, 430072,

Wuhan, China

{ogenius, 371234962, 691319000}@qq.com

**Abstract.** With the emerging of UHF RFID, people pay more and more attention to the recognition and selection of multi-tags. As for the recognition problem, it has already been relatively perfectly solved by anti-collision method. But we still lack mature methods to recognize special tags. This article summarizes and discusses two available solutions: one is to adopt accumulative total count distinguish method; the other is to make use of Low-frequency sensei technology.

**Keywords:** UHF RFID, Low-frequency awakening technology, low power consumption, ASH receiving circuit.

## 1 Introduction

RFID is a technology which is developing rapidly in recent years. It has been applied to widespread fields too. It is much more convenient than many other contact identification systems. According to the frequency band, the RFID system can be divided into low frequency, high frequency and UHF. They all have their own applications. Such as supermarkets' fast registration for commodities, personal identification and consumption card in campus, electronic entrance guard, logistic and warehouse management, intelligent parking management, production line management and product testing. In addition, they have also been developed in anti-counterfeiting toll fee, museum collection introduction and vehicle automatic weighing detection. As the requirement for data transmission rate and identification distance has been improved, UHF RFID has been among the first to be developed in many application fields.

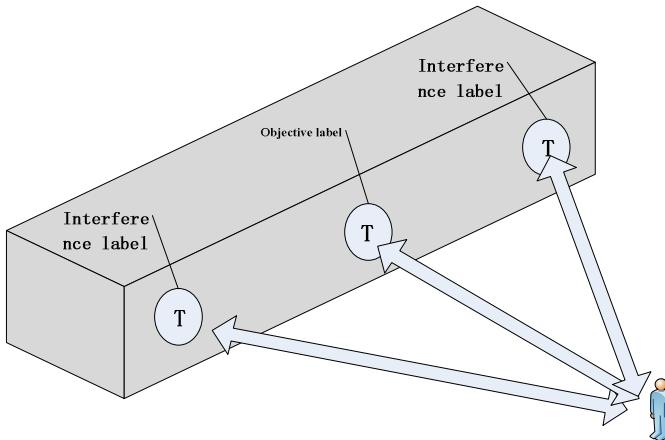
In the application of UHF RFID system, we often encounter this kind of problem: though using prevent collision algorithm can relatively perfectly identify the multiple tags of reader's readable scope, sometimes we don't need the information of all labels, but the information of a tag in a particular position. For example, when users only need the information of the nearest label within a card reader's readable range, the information that upper computer is to process only relies on data directly or indirectly obtained by the card reader, and simple recognition system is unable to provide it. The information unscreened cannot be the final object. Hence, this requires some new methods.

Because the general reader uses directional antenna, only when electronic tags are in the reader's range of the directional antenna can they be identified, and then next operation is executed. However, due to the increase of some applications' request to

reading distance, it is more likely to appear several electronic tags in identified areas at the same time. In these applications, we firstly need to identify a certain tag. But the other tags will also interfere with the identification process. This relates to antenna arrangement, gain, the direction of the wave graph, antenna lobe width, the reader's launch power, the receiver's sensitivity and RFID tags' position and orientation.

## 2 Solution

In the room shown in Figure 1, when a person stands in a fixed position with a portable reader in hand and there are several labels in front of him, we need to attain the information from a nearest tag. In order to overcome the problem that a reader can only "recognize" the specific tag but cannot "distinguish" it, we can adopt the following solutions.



**Fig. 1.** The scenario diagram of the tags that need to be distinguished

### 2.1 Cumulative Frequency Distinguish Method

When there is a relatively long distance between the operator and interference label, objective label, we can distinguish them with the method of comparing the readable cumulative frequency in unit time. That is, when we need the information of the frontal label, we start literacy device to make it read fast and continuously. After a period of time, we compare the respectively accumulative total of all the labels read in. We select the biggest accumulative total number and consider it as "a different" special label which is just the one in front of the person, so as to distinguish the individual.

With this method we can get what is expected to some extent. However, it has big limitations. When we start literacy device, if the distance is not long enough between each label, the response frequency in unit time will be almost the same, which will lead to the mistake of information screening. But this method doesn't have too many

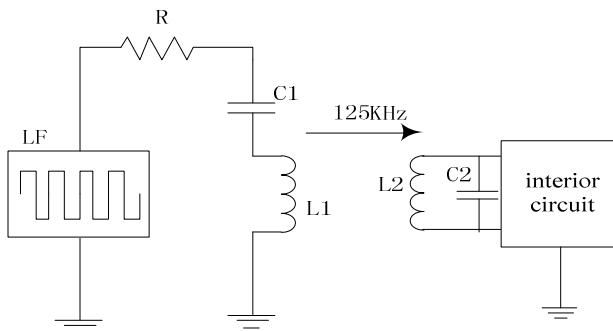
requires in hardware of RFID system, so we can save a lot of cost. At the same time, when the label changes dynamically as well as its distribution is almost in sequence, and the literacy device is fixed, this kind of method is more suitable.

## 2.2 Implementation Methods of Low-Frequency Awakening Technology

**Active Tags Power Consumption Problems.** Low-frequency awakening technology comes from LC oscillating circuit. As is shown in figure 2, it is mainly composed of a low-frequency oscillation circuit which consists of L1 and C1 and parallel resonant circuit formed by L2 and C2 [2]. The inherent frequency of the circuit is

$$f = \frac{1}{2\pi\sqrt{LC}} \quad (1)$$

When the antenna takes over signals of inherent frequency, the circuit will resonate and excite the biggest induced EMF [3]. The awakening distance can be changed with the frequency.



**Fig. 2.** Schematic of the Low-frequency awakening circuit

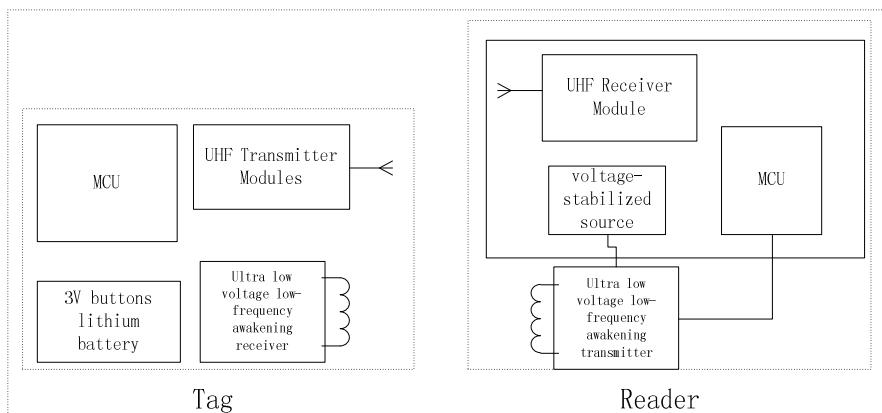
This scheme adopts Active tags. Low-frequency awakening technology can solve the problem of power consumption and maintenance more effectively compared with timing awakening technology of active tags. In power saving mode, all the circuits in electronic tag circuits waked up by LF awaking technology are in the dormant state. At the same time its Clock circuit doesn't work and has no power consumption. But the timing awakening circuit has the problem that clock circuit is always in working mode no matter whether the labels are working or not. Its MCU can't go into the dormant state. So Low-frequency sensei method can make the working life of labels much longer, and so it is with the cycle of System maintenance.

**Avoid the Influence of Interference Tags.** We can avoid the influence of interference tags in high-frequency communication way. By changing the power and frequency of low-frequency transmitter, we can control its waken-up distance within the scope which we need (waken-up distance should be shorter than UHF RFID). First,

low-frequency antenna, in the card reader end, sends out low-frequency signals as the interrupt signal to RFID. The interrupt signal wakes up processor circuit of RFID to switch tags from sleep mode to operative mode. Then we can operate on the tags. In this scheme, distances between tags under operation are controlled within the distance that can be waked by low frequency, thus the amount of tags which is likely to cause interference is reduced.

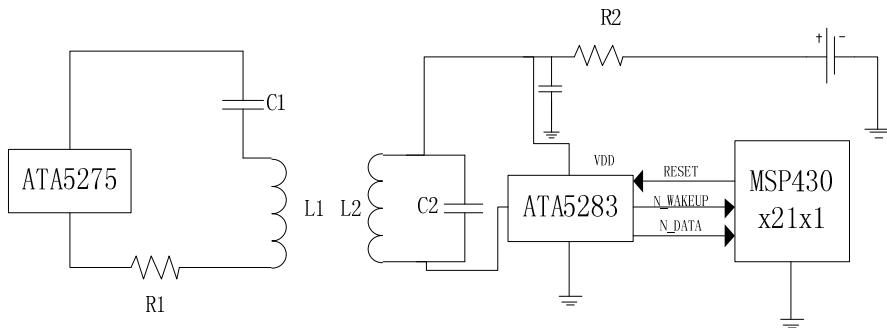
### 2.3 The Design of Hardware Circuit

UHF FRID with low-frequency awakening technology can be designed as is shown in figure 3. Low-frequency receiver, MCU and high-frequency transmitter are all in sleep mode when idle. The ultra low voltage receiver chip will be waked when it receives the message from the low frequency transmitter of the reader. Then it will trigger MCU at N\_WAKEUP pin by producing a low level. Consequently high-frequency devices will be triggered for communications of high-frequency data.



**Fig. 3.** The UHF FRID system applied low-frequency awakening technology

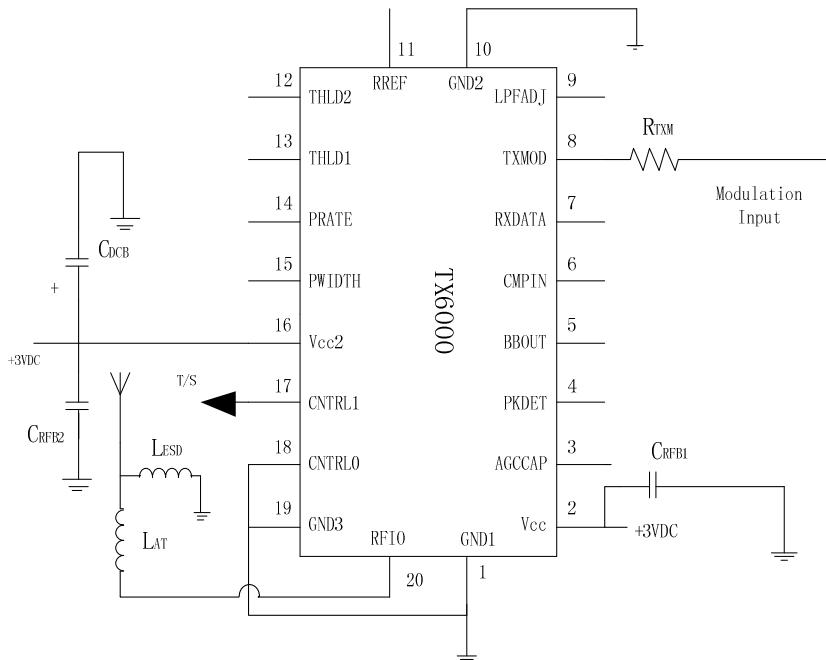
**The Design of Low-Frequency Awakening Circuit.** As for the design of low-frequency awakening circuit, we can use ultra low-voltage awakening transmit chip (ATA5276) and receive chip (ATA5283) produced by Atmel. The current of ATA5283 in operating mode is  $2\mu\text{A}$  and the current in standby mode (listen mode) is  $0.5 \mu\text{A}$ . The low power property makes it very suitable for active RFID. Its operating voltage ranges from  $2\text{V}$  to  $3.6\text{V}$  [4]. As for the MCU of the RFID, we can choose 16-bit ultra-low power consumption mixed-signal processor MSP430x21x1 series which is provided with RISC, whose operating voltage ranges from  $1.8$  to  $3.6\text{V}$ , and the current in active mode is  $250\mu\text{A}$  while in standby mode is only  $0.7\mu\text{A}$ . It costs less than  $1\mu\text{s}$  waking from standby mode, and operating ambient temperature is  $-40\sim+85^\circ\text{C}$ . LF awakening circuit can be designed as is shown in Figure 4.



**Fig. 4.** Structure of the LF awakening tags

### Design of High Frequency Communication Circuits

*Design of High-frequency transmitter circuit.* We use RFM's TX6000 transmitter chip in the part of high frequency communication, whose transmit frequency band is from 916.3 to 916.7MHz. This chip can use the Modulation Mode of ASK/OOK. Its maximum transmission-rate of the RF data is 115.2Kbps, and the operating current is 12mA in the transmit mode while the operating current is 0.7 $\mu$ A, operating voltage 2.2-3.7V, operating temperature -40~+85 °C under the sleep mode. Transmitter circuit modulated by ASK can be designed as is shown in Figure 5, the modulated data can be accessed from the TXMOD pin.



**Fig. 5.** High-frequency transmitting circuit

*Design of High-frequency receiving circuit.* To overcome the shortcomings that the requirements of super heterodyne receiver for antenna impedance matching are more demanding, we can use RX6000 receiver chips supporting TX6000 to compose ASH (Amplifier-Sequenced Hybrid) receiver circuit which can stably work at a wide antenna impedance range. RX6000's working voltage is about 2.2-3.7V, and working current is  $0.7\mu A$  in sleep mode. In receiving mode, working current will be 3.0mA when data transfer rate is 2.4kbps (RPR=330K); working current will be 3.1A when it reaches 19.2kbps (RPR=330K). When high data transfer rate is 115.2kpbs, working current will be 3.8mA.

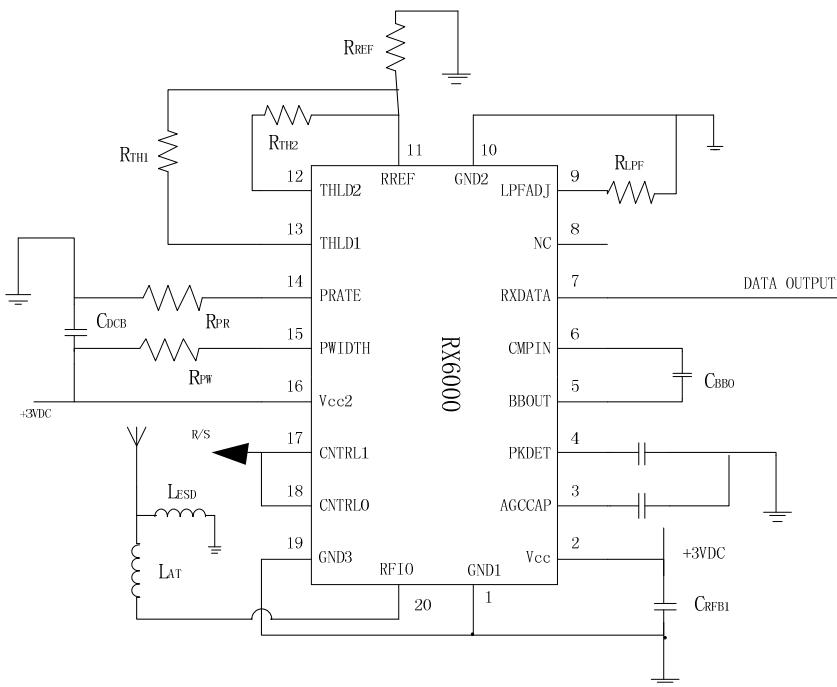


Fig. 6. High-frequency receiving circuit design

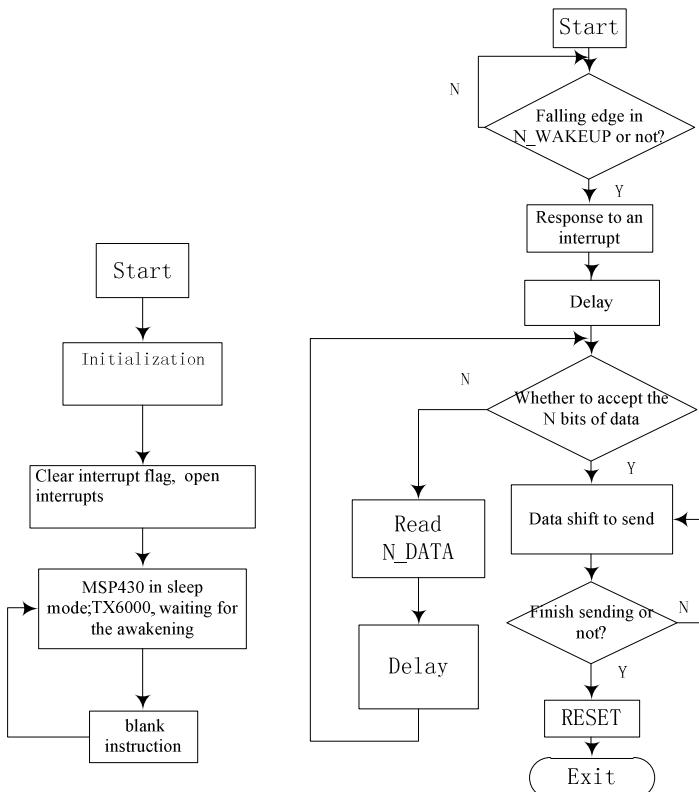
In UHF link application, ASH receiver circuit has high stability. ASH also has strong channel capture property that only the strongest signal in the UHF occasion can play a leading role in receiving circuit, and weaker signals will be ignored by the receiving circuit. This property can prevent Co-Channel interference effectively. Receiving circuit design is shown in Figure 6.

## 2.4 Process Flow Design

When there is a signal on the falling edge, ATA5238 will trigger MSP430 to produce interrupt response, then start interrupt service routine in which data reading will be

completed. After that, MSP430 will start the high-frequency radiating circuit to send shifting data. When data sending is finished, MSP430 sends a positive pulse RESET to ATA5283 to make it re-enter into standby mode. MSP430 turns into sleeping mode after the exiting interrupt service routine, at the same time TX6000 also becomes dormant.

Tag main program and interrupt program flow chart is shown in figure 7.



**Fig. 7.** Tag main program and interrupt program flow chart

### 3 Summary

These two schemes are used in different situations. The former is suitable for the case that there is a long distance and relative motion between tags and the reader, and investment is limited. The latter applies to static environment. Under the premise of not to lose too much cost performance, the introduction of low-frequency awakening technology solves two problems. Firstly, it solves the problem of the power consumption and life issues of active tags. Secondly, it effectively reduces the possible label conflict and makes up the defects of low data transmission rate and small information capacity in low-frequency systems together with making use of the characteristics that high-frequency can be loaded with large information and data transmission rate is high.

## References

1. Xiao, W.: Neighbor disturbance solution in RFID application. Automatic Identification Technology of China, 46–47 (2007)
2. Weiming, L.: Radio frequency identification (RFID) technology principle and application. China Machine Press, Beijing (2006)
3. Shihua, C., Fang, Z.: Application and research in micro-consumption electronic active RFID using the low-frequency sensei technology. Control and Automation Publication Group, 231–234 (2008)
4. Atmel. Interface IC for 125 kHz Wake-up Function ATA5283(EB/OL),  
[http://www.icpdf.com/PdfView.asp?id=93605\\_679896](http://www.icpdf.com/PdfView.asp?id=93605_679896)
5. Zhiming, A., Sun, S., Wei, K., Liu, Z.: Design and application of passive keyless vehicle entry system. Application of Electronic Technique, 48–51 (2007)

# Sensor Ontology Building in Semantic Sensor Web

Yimin Shi, Guanyu Li, Xiaoping Zhou, and Xianzhong Zhang

Information Science and Technology College,

Dalian Maritime University, Dalian, China

shiyimin1966@126.com,

{rabtitlee, zxp163, jiutianyilong}@163.com

**Abstract.** To address the shortcomings of existing sensor ontologies, such as lack of a unified ontology framework and being built completely manually, this work proposes the sensor ontology framework and algorithms for automatic sensor ontology update and extension. In addition, sensor discovery based on time, space and theme is implemented. Sensor core ontology and its update and extension are evaluated. The experimental result reveals that the proposed sensor ontology framework is feasible, and the algorithms for sensor ontology update and extension are effective.

**Keywords:** Sensor Ontology, Sensor Core Ontology Building, Sensor Ontology Update and Extension, Semantic Sensor Web, Sensor Discovery.

## 1 Introduction

To overcome the lack of semantic in sensor networks and sensor web, Amit Sheth et al. presented Semantic Sensor Web (SSW) in 2008 [1], which leverages current standardization efforts of the Open Geospatial Consortium (OGC) and Semantic Web Activity of the World Wide Web Consortium (W3C). The purpose of this work is to provide a conceptual model for sensor data, and thus accomplish knowledge reasoning as well as sharing of sensor data by sensor ontology.

Compton et al. [2] provided a survey of the state of the art of sensor ontologies. Typical sensor ontologies were reviewed and analyzed for the range and expressive power of their concepts, and their reasoning ability as well.

The current sensor ontologies have the following shortcomings: (1) Lack of a unified ontology framework and consistency in definition of concepts, which leads to poor sharing and reuse; (2) Lack of explicit hierarchy and related poor logic expressiveness, resulting in unsatisfactory reasoning; (3) Being built completely manually, making it impossible to implement update and extension automatically, and making the ontologies small with low building efficiency.

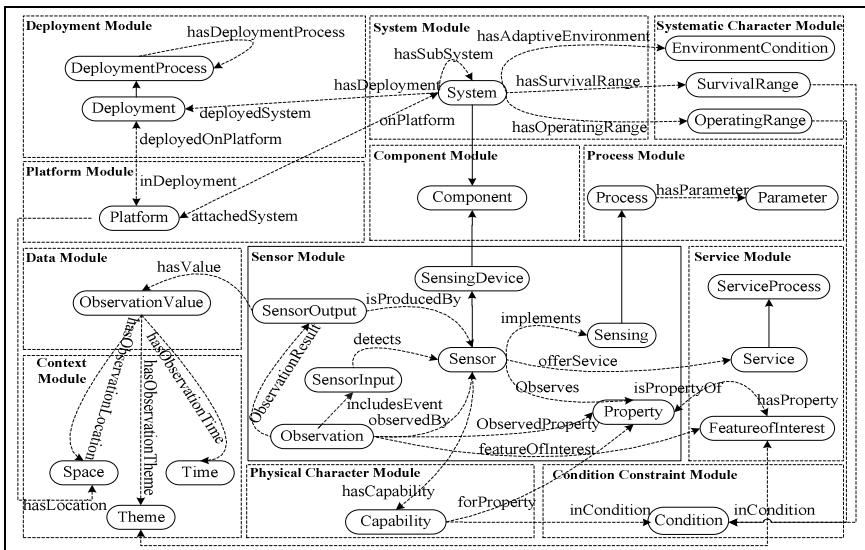
On the basis of SSW, the framework of sensor ontology is proposed according to properties and observations of sensor, and sensor core ontology (SCO) is built, then automatic update and extension of sensor ontology are realized.

## 2 Sensor Ontology Building

### 2.1 SCO Building

The approach for SCO building is to reuse concepts of current ontologies and commonly used concepts of sensor domain, and adopt instances of weather sensor from the Internet, and a data set provided by Kno.e.sis Center of Wright State University, which contains hundreds of thousand observations and related information collected by weather sensors of the United States weather station during the years from 2001 to 2008.

The sensor ontology framework is separated into twelve modules. The framework and the main concepts and their relationship are shown in Fig. 1. The Sensor Module is the core module of the entire sensor ontology model, and it serves as the bridge connecting the other modules. Sensor, Process, System and Component are the most basic concepts. The concept Sensor in the Sensor Module stands for the sensor devices for measuring. The concept Process, a part of the Processing Module, stands for the processing of input and output sensor signals. The concept System in the System Module stands for mixed sensing devices containing software programs, sensors, collectors and power. The concept Component is the parent of all the classes. It describes information of all man-made objects and belongs to the Component Module. There are additional twenty two main concepts with their thirty-three relations, which are also depicted in Fig. 1.



**Fig. 1.** Sensor Ontology Framework

The implementation of SCO is divided into three stages: Definition of class and class hierarchy by referencing existing sensor domain ontologies and top ontologies such as SUMO and DULCE, and also national sensor standards; definition of properties and facets of properties; coding of SCO.

## 2.2 Sensor Ontology Update and Extension

Based on ontology learning framework [3, 4], instance update and concept extension of sensor ontology are implemented and then the sensor ontology with domain character and integrity is generated; therefore, explicit and coincident sensor ontology is achieved by modifying and improving.

The process of ontology update is to add individuals in candidate concept set into SCO to enrich sensor ontology. By imitating sensor observation, real-time data for instance update of the class “Observation Value” in sensor ontology is provided to enable the instance update of classes as Time, Space and Theme in sensor ontology, and the relation assertions of time, space, theme and observation are driven.

The process of ontology extension is to match and extend the classes in SCO using the concepts in candidate concept set. Firstly, the correlation between sensor ontology concepts and domain documents is calculated by using domain correlation [5], and then the documents with low correlation are deleted. Secondly, edit distance similarity [6] and context similarity [7] between a concept in sensor ontology and a term in the domain document are calculated. Lastly, the two similarities are integrated using the Sigmoid function as a coordinate value, named concept matching degree, and judgment condition for concept matching is obtained.

Given two terms  $t_1$  and  $t_2$ , their edit distance similarity noted as  $\text{Sim}_{\text{EditDist}}(t_1, t_2)$  is given as formula (1).

$$\text{Sim}_{\text{EditDist}}(t_1, t_2) = 1 - \frac{\text{EditDist}(t_1, t_2)}{\text{Max}(\text{len}(t_1), \text{len}(t_2))} \quad (1)$$

Where  $\text{EditDist}(t_1, t_2)$  stands for the edit distance between the two terms  $t_1$  and  $t_2$ , and  $\text{len}(t_1)$  and  $\text{len}(t_2)$  stand for the string length of  $t_1$ , and  $t_2$  respectively.

Setting the context term expression of  $t_1$  with  $\text{Context}_{t_1} = \{t_3, t_5, t_6, t_7\}$ , and the context term expression of  $t_2$  as  $\text{Context}_{t_2} = \{t_3, t_4, t_5, t_7\}$ , their vectorial representations then are shown as follows:  $\vec{W}_{t_1} = \{t_3/w_3, t_5/w_5, t_6/w_6, t_7/w_7\}$ ,  $\vec{W}_{t_2} = \{t_3/v_3, t_4/v_4, t_5/v_5, t_7/v_7\}$ ,  $w_i$  and  $v_j$  stand for weight of context term.  $\vec{W}'_{t_1}$  is the extension of  $\vec{W}_{t_1}$ , and so is  $\vec{W}'_{t_2}$  to  $\vec{W}_{t_2}$ ,  $\vec{W}'_{t_1} = \{t_3/w_3, t_4/0, t_5/w_5, t_6/w_6, t_7/w_7\}$ ,  $\vec{W}'_{t_2} = \{t_3/v_3, t_4/v_4, t_5/v_5, t_6/0, t_7/v_7\}$ . The context similarity between  $t_1$  and  $t_2$  expressed as  $\text{Sim}_{\text{Context}}(t_1, t_2)$  is given as formula (2).

$$\text{Sim}_{\text{Context}}(t_1, t_2) = \cos(\vec{W}_{t_1}, \vec{W}_{t_2}) = (\vec{W}_{t_1} \cdot \vec{W}_{t_2}) / \sqrt{|\vec{W}_{t_1}|^2 + |\vec{W}_{t_2}|^2} \quad (2)$$

The calculation of  $\vec{W}'_{t_1}$  and  $\vec{W}'_{t_2}$  is as follows: For the term  $t_i$ , if  $t_i \in \text{Context}_{t_1}$ , and  $t_i \in \text{Context}_{t_2}$ , then do nothing; if  $t_i \notin \text{Context}_{t_1}$  and  $t_i \in \text{Context}_{t_2}$ , then append  $\vec{W}'_{t_2}$  with  $t_i$ , and set the weight of  $t_i$  zero; if  $t_i \in \text{Context}_{t_1}$  and  $t_i \notin \text{Context}_{t_2}$ , then append  $\vec{W}'_{t_1}$  with  $t_i$ , and set the weight of  $t_i$  zero.

The concept matching degree noted as  $\text{Sim}_{\text{Concept}}(t_1, t_2)$  is given as formula (3).

$$\text{Sim}_{\text{Concept}}(t_1, t_2) = \alpha \text{Sim}_{\text{EditDist}}(t_1, t_2) + \beta \text{Sim}_{\text{Context}}(t_1, t_2) \quad (3)$$

Where  $\alpha$  and  $\beta$  are different coordinate factors and they take the form of the Sigmoid function.

### 3 Application of Sensor Ontology in SSW

#### 3.1 Design of SSW System

To verify the feasibility and effectiveness of Fig. 1 and algorithms for sensor ontology update and extension, a prototype system named Sensor Ontology-Driven Semantic Sensor Web (SenOntDSSW) is designed, which implements the sensor discovery after the update and extension for sensor ontology has been finished.

The architecture of system SenOntDSSW is show in Fig. 2. In system SenOntDSSW, five main modules are included, which respectively are preprocessing, sensor ontology update, sensor ontology extension, sensor ontology improvement and sensor data query. Sensor ontology update module and sensor ontology extension module are the core modules of this system.

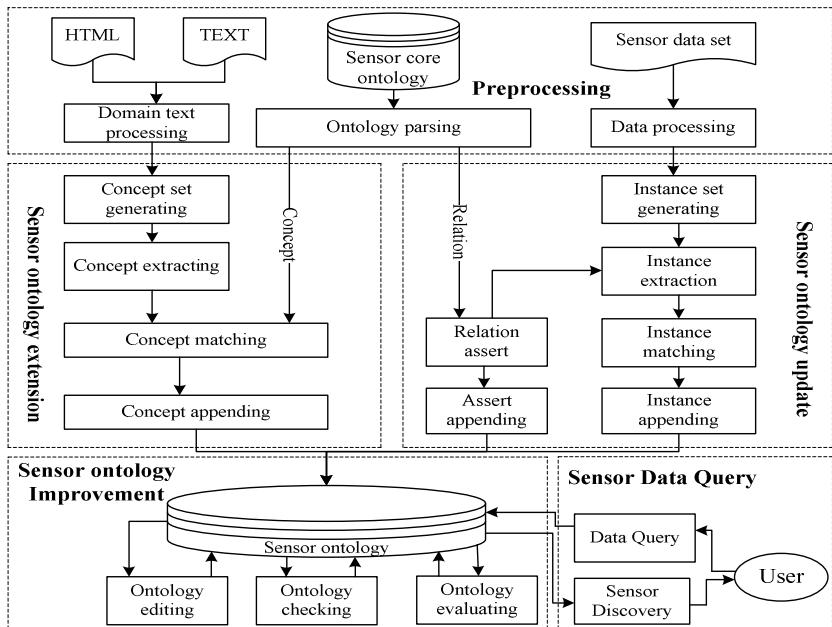


Fig. 2. Architecture of SenOntDSSW

#### 3.2 Sensor Discovery

Sensor discovery means, when facing large amount of sensor data in the web, sensor data “consumers” can get their interested sensors or sensor observation easily, and at the same time, the sensor data “producer” can clearly know the status of data use, such as usage location, time, and the users.

The sensor characters of space, time and theme is described with sensor ontology, and semantic querying about these is implemented, so that the eligible sensor will be found which can provide the users with the information of the sensors and their observations.

This paper focuses on the semantics of sensor data, instead of sensor data collection or transformation. So in the experiment, we assume that sensor data is stored in local servers. The experimental procedures are shown as following:

**Sensor Simulative Deployment.** Campus of DLMU is set as the sensor simulative deployment region.

**The Semantic Annotation of Time, Space and Theme.** New semantic information of time, space, and theme are added for deployed sensors. The time, space, and theme semantics of sensor data represented with OWL2 is shown in Table 1. The information there means that at the time of 2011-04-18 09:30:00, the temperature of the air measured by the A301 sensor located in DLMU is 12°C.

**Provenance Query of Sensor:** It means distinguishing sensor and analyzing observation by time and space information of data entity, and returning all provenance information of the data. Firstly, input “2011-4-18\_9:30:00”, then the system transforms it to SPARQL query, thus the sensors that satisfy the query condition is obtained. When the sensor is clicked, the information of the sensor is acquired.

**Table 1.** Semantic Annotation of Sensor Ontology

---

*owl:<http://knoesis.wright.edu/ssw/ont/sensor-observation.owl#>*

---

```

1: <owl:NamedIndividual rdf:about="#A301">
2:   <rdf:type rdf:resource="#Sensor"/>
3:   <hasObservationValue rdf:datatype="&xsd:string">
4:     12 °C</hasObservationValue>
5:   <hasResponseTime
6:     rdf:resource="#2011-4-18_9:30:00"/>
7:   <hasLocalLocation rdf:resource="#DalianMaritimeUniversity"/>
8:   <hasObservationTheme rdf:resource="#AirTempreature"/>
9: </owl:NamedIndividual>
10: <owl:Class rdf:about="#Location"/>
11: <owl:Class rdf:about="#Sensor"/>
12: <owl:Class rdf:about="#Theme"/>
13: <owl:Class rdf:about="#Time"/>

```

---

The results of the system demonstration show that the query based on the context information as time, space, and theme can be implemented when sensor ontology is introduced into SSW. Besides, different sensors will be found depending on different context information, thereby verifying the effectiveness of our sensor ontology.

## 4 Evaluation of Sensor Ontology

### 4.1 Evaluation Results and Analysis of SCO

At present, there is no well-established metrics and tools for ontology evaluation. This paper evaluates sensor ontology objectively by the objective metrics in literature [8]. SCO will be compared with four current typical sensor ontologies: OntoSensor, CESN, OOSTethys and CSIRO. The objective evaluation is given as formula (4).

$$\text{Objective} = I \times w_i + C \times w_c + O \times w_o + P \times w_p \quad (4)$$

$I$  denotes translatability,  $I=N / T$ ,  $T$  denotes total number of terms in ontology;  $N$  denotes the number of terms which can be found in WordNet.  $C$  is clarity,  $C = (\sum 1/A_i) / N$ .  $A_i$  denotes the number of meanings of every translatable term in WordNet, then clarity of each term is  $1/A_i$ ,  $O$  denotes comprehensiveness,  $O=T/M$ .  $M$  denotes the number of terms in standard term set in the domain that the ontology belongs to.  $P$  denotes popularity,  $P=E/H$ ,  $E$  denotes the number of access of the ontology, and  $H$  is the total number of access of all the ontologies in the same domain.  $w_i$ ,  $w_c$ ,  $w_o$ ,  $w_p$  denote the weight of translatability, clarity, comprehensiveness, popularity respectively, and these weights satisfy the equation  $w_i + w_c + w_o + w_p = 1$ .

The data for test of objective evaluation comes from WordNet, literature citation index is used as popularity. The translatability and comprehensiveness of SCO are the highest among the five ontologies, clarity of SCO is medium, popularity of SCO is the lowest while popularity of OntoSensor is the highest. The objective values shown in Fig. 3 are obtained. It can be seen that our SCO receives a better evaluation; it is clearly superior to other ontologies except for OntoSensor.

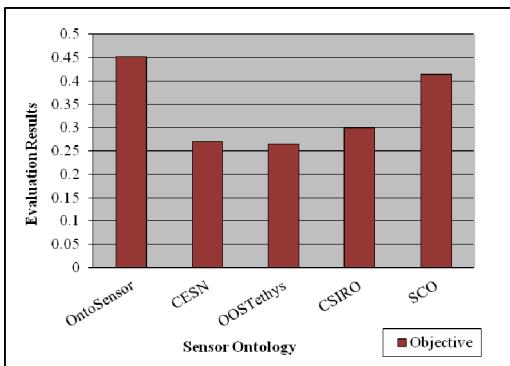


Fig. 3. Objective Review Result

### 4.2 Evaluation Results and Analysis of Ontology Update

An evaluation test of the sensor ontology update implementation has been carried out. The data for the test comes from the weather sensor data set of Wright State University mentioned in section 2.1.

10 sensor data documents are imported into our system, one document at a time. The first five documents do not have observations, while the second half does. The ontology can be loaded to Protégé successfully each time without instance conflict. This illustrates that the updated sensor maintains consistency. The number of updated instances is not necessarily the same even if the number of instances in documents is the same,

because some updated instances already exist in ontology, they do not need to be added to the ontology.

### 4.3 Evaluation Results and Analysis of Ontology Extension

For evaluating concept matching result, precision rate, recall rate and the F-measure metrics are used respectively. The F-measure indicator combines precision and recall rates, as specified in formula (5).

$$F = 2 \times P \times R / (P + R) \quad (5)$$

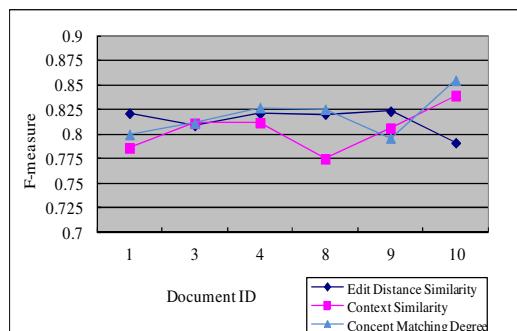
The term documents of the sensor domain come from Web Pages. Near to 1,000 domain terms are chosen from 10 documents with approximately equal numbers of terms, and the domain correlation and concept matching degree are calculated respectively, then the results of document selection and concept matching are analyzed.

The experiment is divided into two stages. The first is domain correlation calculation. The steps are as follows: (1) Determine document selection threshold (We take it as 0.5 here), (2) Calculate the number of terms extracted from each document by concept matching, and (3) Remove the documents with low degree of correlation. The degree of correlation of the removed documents numbers 2, 5, 6, 7 is lower than that of those which are selected. This appears to suggest that the degree of correlation of a document with sensor ontology can affect the matching result of domain terms with ontology.

The second stage is to calculate several assessment measures, including the numbers of both correctly and incorrectly matched terms, as well as omitted terms of edit distance similarity, context similarity and concept matching; and precision, recall rate and the F-measure.

The algorithm for edit distance has a lower precision but higher recall rate. Its F-measure distributes mainly between 0.81 and 0.823 (83 percent), with the highest being 0.823, and the lowest 0.792. The algorithm for context similarity is just the reverse: It has a higher precision but lower recall rate. Its F-measure distributes primarily between 0.786 and 0.811 (67 percent), with the highest at 0.839, and the lowest 0.775. The precision of the concept matching algorithm is higher than that of the edit distance algorithm, and its recall rate is higher than that of context similarity. Its F-measure fluctuates in between, as shown in Fig. 4.

Through harmonizing the two algorithms, the recall rate has been improved effectively while maintaining a higher precision; therefore a better matching effect has been achieved.



**Fig. 4.** Comparison of F-measure

## 5 Conclusions

SCO is built based on existing sensor ontology and knowledge of sensor domain. Sensor ontology update and extension have been implemented, and sensor discovery based on time, space and theme is achieved. The effectiveness of SCO and its update and extension have been verified, and the performance of SCO and its update and extension have been assessed. However, in semi-automatic building of sensor ontology, only concept extension, instance update and relation asserting were realized. Relation learning, rule learning and level extension of sensor ontology should be studied in our future work.

**Acknowledgements.** This work was supported by the National Natural Science Foundation of China under Grant (No. 60972090).

## References

1. Sheth, A., Henson, C., Sahoo, S.S.: Semantic Sensor Web. *IEEE Internet Computing* 12(4), 78–83 (2008)
2. Compton, M., Henson, C., Lefort, L., et al.: A Survey of the Semantic Specification of Sensors. In: 2nd International Workshop on Semantic Sensor Networks, Rome, pp. 17–32 (2009)
3. Du, X., Li, M., Wang, S.: A Survey on Ontology Learning Research. *Journal of Software* 17(9), 1837–1847 (2006) (in Chinese)
4. Fu, K.: The Study of Ontology Learning from Web Pages. Wuhan University of Technology, Wuhan (2007) (in Chinese)
5. Velardi, P., Fabriani, P., Missikoff, M.: Using Text Processing Techniques to Automatically Enrich a Domain Ontology. In: 1st International Conference on Formal Ontology in Information Systems, pp. 270–284. ACM Press, New York (2001)
6. Ristad, E.S., Yianilos, P.N.: Learning String-Edit Distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 5(20), 522–532 (1988)
7. Zhai, L.: Research and Implementation on Semi-Automatic Domain Ontology Acquisition Method. Southeast University, Nanjing (2005) (in Chinese)
8. Ma, X.: The Research and Implementation of a Web-based Ontology Evaluation System. Ocean University of China, Qingdao (2009) (in Chinese)

# Panoramic CIM Model of Power Equipment at Converter Station Based on IOT

Jingyu Huang, Chun Huang, Xiaoqing Huang, Junyong Zhang, and Jie He

College of Electrical and Information Engineering  
Hunan University  
Changsha 410082, China  
nash263@sina.com, zhjy163@yeah.net

**Abstract.** The Internet of Things (IOT), as perception terminals of lifecycle management on power equipment, will undoubtedly be important technical means to facilitate the development of the smart grid. However, traditional information model is not unified that the data and resource can not be shared or used in the monitoring and management of equipment at converter station, which based on IOT. In this paper, the panoramic information model of equipment at converter station based on IOT is explored. Furthermore, we expend the information classes and make the static relationships (Inheritance, Association, and Aggregation) that exist among these classes based on CIM under IOT architecture.

**Keywords:** IOT, panoramic information, CIM, power converter equipment, IEC standard.

## 1 Introduction

The definition of IOT is "intelligent objects or animals" or "smart nodes" such as various monitoring sensors, mobile devices, video surveillance systems, objects accompanied by RFID, with wireless terminal achieving interoperability by a variety of wireless or cable, long distance or short-range communication networks. Its ultimate goal is a kind of integration, whose function is making the management, control of the myriad of objects on earth efficient, safety and environmental [1]. In short, IOT refers to accordance with the appointed agreement having any objects connecting with internet to exchange information and communicate to realize intelligent recognition, orientation tracing, monitor and management through the information peripheral equipments [2]. The SGCC (State Grid Corporation of China) proposed requirement about many core technologies, one of which is the application of IOT Technology in power systems.

The operation of equipment at converter station is important to the reliability of DC transmission, even the safety of power system. The converter equipment have various species, which include converter transformer, smoothing reactor, DC arrester, DC inductor, DC switch, converter valve, etc. Obviously, the composition of information is very complex. Currently, monitoring technology has made remarkable

progress, the state monitoring information of equipment are, however, used inefficiently. This mainly resulted from two reasons: one is devices' lack of comprehensive status information. The other is "Information Isolated Islands" existed in management systems [3]. Thus, we can collect, for instance, nameplate, status, environment and test of power converter equipment in great detail by using technology of IOT, such as sensor technology [4], RFID [5], RTLS [6] and intelligence embedded technology. These technologies have been researched by lots of scholars and engineers. The researches indicate that asset management level of electric power will be improved by IOT Technology. From this point, traditional CIM model is no longer applicable for the converter equipment monitoring based on technology of IOT.

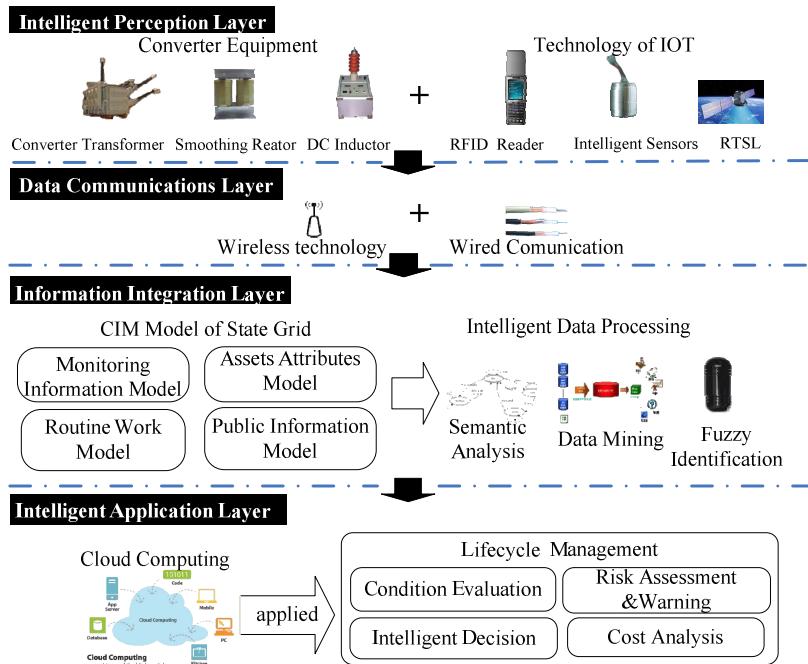
As we know, Converter equipment online monitoring system, dispatching system, manufacturing management information system are respectively applied in different fields. Information of them are inextricably linked, even some of them are exactly the same. Such scenario urges for standardization of inter and intra processes communication and data exchange. At present, there are many integration strategy used by the Electric power enterprise, including data warehouse technology based on CIM/XML, integration based on UIB and technology of integration based on SOA [7]. Whichever integration strategy is used, without question, it's executed by the advanced computer technology with CIM model. Therefore, panoramic information model of power converter equipment is desperately needed.

This paper extends the IEC61968, IEC61970 and some other relative international standards to develop a panoramic CIM model of converter equipment based on the technology of IOT by using object-oriented techniques, which could meet the requirements of SGCC application system.

The rest of the paper is organized as follows: in Section 2 we present the proposed architecture of IOT in information model while in Section 3 we give the classification of information considering RFID. Section 4 presents the CIM model of converter equipment by tool of Rational Rose and Section 5 brings conclusions of this paper.

## 2 Architecture of IOT in Converter Equipment CIM Model

Asset Life-cycle Management is a kind of management methods that considering overall process of assets that includes planning, designing, construction, acquisition, operation, maintenance, and abandonment. Its purpose is to make the equipment life-cycle cost minimum when the efficiency is guaranteed. By means of technology of IOT, for example, collecting panoramic status information by various sensors which used to evaluate equipment status, or predict the lifetime, is more comprehensive. Hence the diagnosis or assessment will be more precise and instantaneous. The Architecture of IOT applied to converter equipment is presented in Figure 1, which is a four-layer structure, i.e., the intelligent perception layer, the data communications layer, the information integration layer and the intelligent application layer.



**Fig. 1.** Architecture of IOT applied to converter equipment

The intelligent perception layer contains various kinds of tags encoded by EPC (Electronic Product Code) and intelligence sensor, RFID reader, cameras, etc. They contribute to the perception of equipment information. The data communications layer provides transparent and resolvable data transmission channel. Monitors can transmit data to municipal power condition monitoring system through the wireless communication network. The information integration layer savant middleware process the EPC read by RFID reader. It also has the function as converging, storage, fusion and delivery by using technology of fuzzy identification, data mining, semantic analysis, etc. The intelligent application level adopted the method of Modular-Design. Large amounts of information can be acquired by distributed large-scale platforms by using Cloud Computing technique [8]. Though the CIM itself involves almost all typical models, extension and improvements of the standard are constantly needed especially in the application layer.

### 3 Information Classification Considering EPC and RFID

On the basis of requirement analysis of the information model, we believe it is necessary to classify converter equipment information. The information mainly can be categorized into four types: equipment fundamental information in RFID tags, asset

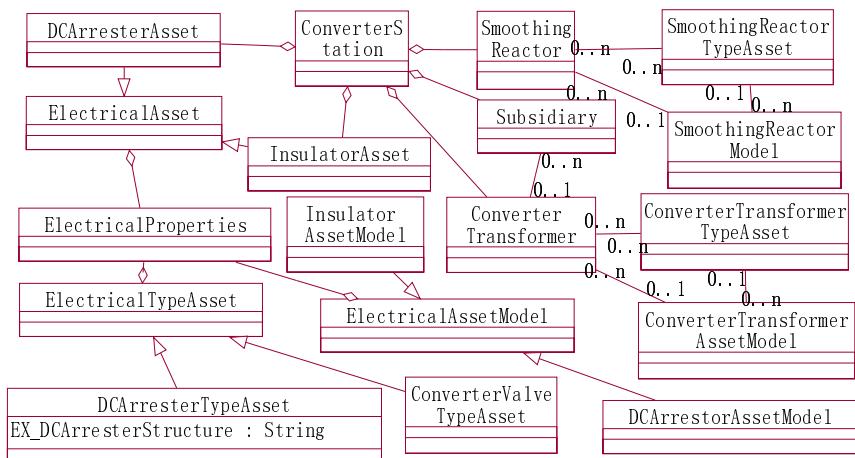
information by manual entry, public information and operation information measured by secondary devices.

The fundamental information in RFID tags which primarily store the invariable or slowly changing data [9]. It supports the nameplate information, technical parameters, code of operating equipment, etc. The adaption of converter equipment information model encoded by EPC to the tradition information model based on IEC standard should be the key consideration. Asset information by manual includes information of preventive test, defects, inspection, assessment, maintenance record and defect eliminating record. Public information composed of weather, public security information and so on. Weather information is the microclimate of monitoring sections. Public security information mainly involves the information of extreme weather, natural disasters, thief alarming, war and other relative message. Operation information reflects the present state of a power system. The public information and operation information needed to be collected in real time.

## 4 Extended Model of Converter Equipment Based on the CIM

### 4.1 Assets Attributes Modeling Based on CIM

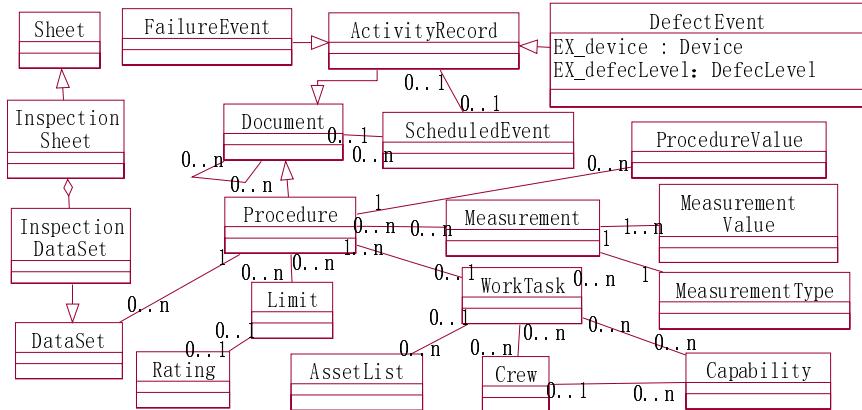
The assets attributes of at converter station are divided into three classes: Asset, TypeAsset and AssetModel. It is noted that converter equipment are not modeled in the existing CIM. Therefore, it is real time we extend their attributes and place them into these three classes, to finish the modeling of assets attributes information, we should correlate these classes, which schematically illustrated in Figure 2 by the Rational Rose tool.



**Fig. 2.** CIM model of assets attributes of converter equipment

## 4.2 Routine Work Modeling Based on CIM

The power converter equipment need regular check-ups that including inspection, test, maintenance and diagnosis. The relative data should be recorded in process of them. As we know, equipment state evaluation is based on information of inspection, operation, etc. The synthetic judgment can not be made before we possess information. As we talked before, the information can be acquired abundantly and instantaneously by the core technique of IOT. Thus, in order to obtain abundant information for the Life-cycle management, modeling of the routine work is extremely essential. The inspection model based on IOT is shown in Figure 3 by Rational Rose.



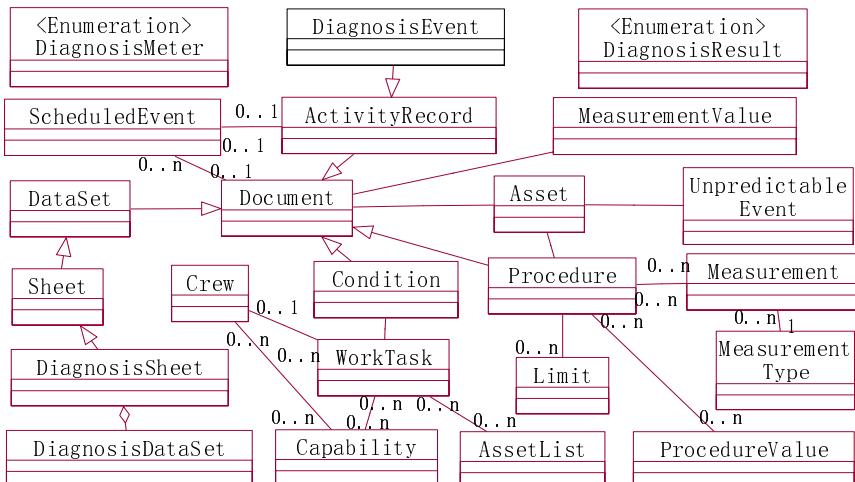
**Fig. 3.** Inspection model of converter equipment

Inspection procedure and measurement are closely linked. The class Measurement represents all the items that should be metered. MeasurementValue is a table that contains the restriction value of these items. Crews are essential to equipment inspection. Every worker is connected with a task which modeled as class WorkTask. Crews can take a handheld RFID reader to get useful information from converter equipment encoded by EPC. The results of inspection are also very important which provide the defect level of equipment, type of the devices and so on. They can also be modeled. They will be translated immediately by wireless technology, or as a form of sheet which recorded in the Sheet class. All the records belong to type of Document. Crews finally find out the class FailureEvent and DefectEvent according to the WorkTask class. The class ScheduledEvent represents a model of the message which specifies time schedule of inspection and identifies path and for which the schedule is specified. The Device class and DefecLevel class, which shown in Figure 3, are the type of enumeration. Figure 4 shows the attributes of these two classes.

<p>&lt;Enumeration&gt;</p> <p>Device</p> <p>converterTransformer smoothingReactor DCArrester DCInductor DCSwitch converterValve protectDevice DCTransmissionLine</p>	<p>&lt;Enumeration&gt;</p> <p>DefectLevel</p> <p>serious moderate mild</p>
--	--

**Fig. 4.** Attributes of device class and DefectLevel Class

With the similar method, the diagnosis model can be established referring to principle of electrical converter equipment diagnosis [10], which illustrated in Figure 5 by Rational Rose. The diagnosis model indicates all the information which used in diagnosing power converter equipment. On the background of IOT Technology, there are many new high-tech equipments in this model which listed the DiagnosisMeter class. For the most part, fault diagnosis method of converter transformer is testing of “oil chromatogram”. Measurement of dissolved gases in the converter transformer oil may produce several attributes, including  $H_2$ ,  $CH_4$ ,  $C_2H_2$ ,  $C_2H_4$ ,  $C_2H_6$ , etc, which are contained in the class Measurement. The class DiagnosisMeter is a profile for the meters used which includes smart gas sensor whose message can be transited by wireless transmission. The class DiagnosisResult may contain a set of different types of readings by element in DiagnosisMeter class. Finally, these meters, results of the diagnosis and some other useful information which is available to the diagnosis, will all belong to the DiagnosisEvent class.

**Fig. 5.** The diagnosis model of converter equipment

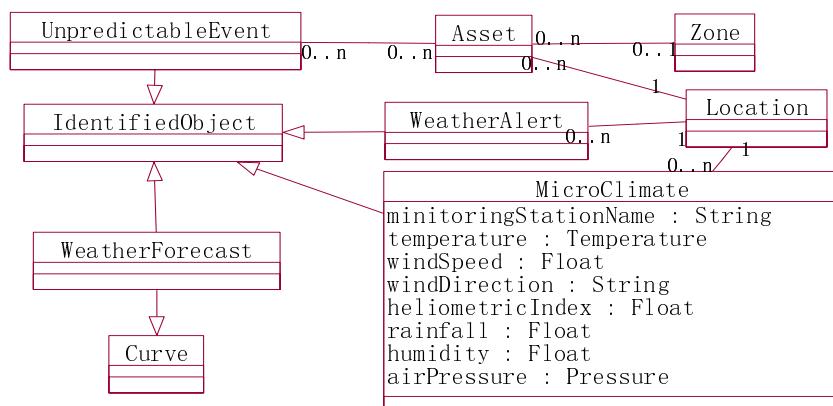
#### 4.3 Monitoring and Public Information Modeling Based on CIM

Converter equipment monitoring is related to features of equipment such as electrical insulation and mechanization. Monitoring parameters consist of electric, machine and some technology based on technology of IOT. From this point, all the information about equipment of converter station monitoring and measurement should be modeled. The class Measurement is used to describe information about the time, reasons of equipment measurement. Measurement results are digitalized in MeasurementValue table according to readings obtained from the meter. MeasurementType models all kinds of applied measurement types.

With technology of IOT, data can be collected through wireless sensor nodes, which connect with Wireless Communication Link to support wireless communication by transmission technology network, such as WiFi, Bluetooth and ZigBee. It sends data to MonitoringTerminalUnit [11]. Data are sent to monitoring center by GSM or GPRS. These classes should be also designed into Monitoring Information Model.

Converter Stations are extensively distributed, even serious the equipment in them. Hence public information, which mainly concerns the environment information and public safety information, has great effect on the safe running of electric converting, even the power network. The weather information and warning of meteorological disaster where the monitoring points located should be modeled. Weather information comes from meteorological department. Alarming information of stealing and war should be taken into consideration. The public information of power converter equipment mainly involves the extended class WeatherAlert, Microclimate, UnpredictableEvent and their static relationships with class Assets and Location. For information of weather forecast is modeled by time-varying curves that the class Weather Forecast inherited the Curve class which has been defined in the IEC61968[12].

The model of public information is shown in Figure 6 using tool of Rational Rose. The attributes of the MicroClimate class are listed. Data formats of each attribute are also included which have been already defined in the Area Package in IEC61970.



**Fig. 6.** Public information model of converter equipment

## 5 Conclusion

Technology of IOT could perceive status information of electric equipment on a large scale quickly. Therefore, it will surely have tremendous impact on lifecycle-management performances of electric equipment. In this paper, the architecture of IOT for the model of converter equipment is established and the panoramic CIM-based model of power converter equipment based on the IOT is explored. This model will realize sharing of converter equipment information resources on the basis of the IOT.

**Acknowledgments.** Supported by the National High Technology Research and Development Program of China (863 Program) (Grant No.2011AA05A120), Scientific Innovation Program of Postgraduate in Hunan (Grant No. CX2011B144).

## References

1. Dabholkar, A., Gokhale, A.: An Approach to Middleware Specialization for Cyber Physical Systems. In: 29th IEEE International Conference on Distributed Computing Systems Workshops, pp. 73–79. IEEE Press, Montreal (2009)
2. Zhou, H.: Internet of Things: Technology Standards and Business Models. Publishing House of Electronics Industry, Beijing (2010) (in Chinese)
3. Zhang, J., Guo, C., Cao, Y.: Substation Equipment Condition Monitoring System and IEC Model Coordination. Automation of Electric Power Systems 20, 67–72 (2009)
4. Urien, P.: HIP-Tags Architecture Implementation for the Internet of Things. In: First Asian Himalayas International Conference on Internet, pp. 1–5. IEEE Press, Katmandu (2009)
5. Yan, B.: Supply Chain Information Transmission Based on RFID and Internet of Things. In: ISECS International Colloquium on Computing, Communication, Control and Management, CCCM, Sanya, pp. 166–169 (2009)
6. Mayordomo, I., Spies, P.: Emerging Technologies and Challenges for the Internet of Things. In: 2011 IEEE 54th International Midwest Symposium on Circuits and Systems, pp. 1–4. IEEE Press, Seoul (2011)
7. Dai, P.: Design and Implementation of ESB Based on SOA in Power System. In: International Conference on Electrical and Control Engineering, Weihai, pp. 519–522 (2011)
8. Shi, W., Liu, M.: Tactics of Handling Data in Internet of Things. In: 2011 IEEE International Conference on Cloud Computing and Intelligence Systems, CCIS, Beijing, pp. 515–517 (2011)
9. Darianian, M., Michael, M.: Smart Home Mobile RFID-Based Internet-Of-Things Systems and Services. In: International Conference on Advanced Computer Theory and Engineering, ICACTE 2008, Phuket, pp. 116–120 (2008)
10. Kim, J.: A Diagnosis Method of DC/DC Converter Aging Based on the Variation of Parasitic. In: 30th Annual Conference of IEEE on Industrial Electronics Society, pp. 3037–3041. IEEE Press, Incheon (2004)
11. Li, S., Xu, L., Wang, X.: Compressed Sensing Signal and Data Acquisition. In: Wireless Sensor Networks and Internet of Things, p. 1. IEEE Press (2012)
12. Lv, G., Liu, H.: Study on IEC 61970/61968 Based Information Integration for Smart Distribution Grid. In: International Conference on Electrical and Control Engineering (ICECE), pp. 5019–5022. ICECENG, Yichang (2011)

# An Infrared Ranging System for Automotive Anti-collision

Liang Xu and Zhiqiang Meng

College of Electrical and Information Engineering,  
Hunan University, Changsha, China  
[xuliang186@126.com](mailto:xuliang186@126.com)

**Abstract.** In this paper, we present a high-frequency current source transmitter circuit, which uses the high-power composite pipe as the regulator and the high-speed op-amp as the feedback link. As matching the appropriate electrical parameters of the current output at high frequencies, the anti-collision system can be maintained in constant state and can be adjusted in real-time based on distance stalls in the larger context. Experimental results shows that the constant current source circuit can achieve the output gear continuously adjustable and the maximum current is up to 500 mA when the input frequency is 150 kHz. The constant current source circuit can be utilized in automotive infrared range system, and has an extremely important practical value.

**Keywords:** Automotive, anti-collision, infrared ranging, high-frequency constant-current source circuit.

## 1 Introduction

With the economic developing and material level increasing in China, people demand more and more cars. However, the increase of vehicle ownership and other objective issues such as climate, road conditions make the frequent of traffic collision accidents occurrence increase. Therefore, the safety of driving will be more and more important.

To solve the safety of driving problem, the fundamental method is to install a device in a moving vehicle which can measure distance and brake automatically in dangerous distance. Automotive anti-collision system is an active safety technology that can measure distance and brake automatically. A high output current and stable work at high frequent input is the key technology of the design of the automotive anti-collision system.

In this paper, we present a circuit with high-frequency driving and constant-current source. This circuit can output high current and work stably with a high-frequency input signal. A large number of results show that the circuit can switch the output continuously and the maximum current is up to 500mA when the input frequency is 150 kHz, it can be utilized in automotive infrared ranging system and has an extremely important practical value.

## 2 System Components

The automotive anti-collision system is mainly composed of infrared receiver, infrared transmitter, turn signal module, brake control and power supply component. According to vehicle speed, CPU uses sub-file method to control the infrared transmitter for detecting echo signal to measure the distance between the vehicle and obstacles. The vehicle speed is controlled according to the distance to achieve the purpose of preventing collision.

The infrared ranging module is the key technology of automotive anti-collision system. When a car is moving, the infrared pulse-modulated waves are sent to the front of the car. As the infrared waves encountering obstacles, such as other vehicles, the reflected waves are received by infrared receiver, and then the results are dealt by CPU. Therefore, the output current of ranging circuit is requested to be stable and should work steadily at high-frequency input signal.

The control mode of the system is detecting the obstacles within the set distance. C8051F MCU, supplemented by the external interface circuit, is adopted as control chip to send control commands according to the speed of the car, and to regulate transmitting power with sub-file .Each file corresponds to a longest detection range. If the speed of the car is detected in the range R-file, then the corresponding current of the I-file drives the infrared transmitter to send waves. No echo signal indicates that there is no obstacle in the range of R-file and continues to injection using the current of the R-file. With the enhancement of vehicle speed, the stalls corresponding to the speed and the emission current also increases.

Receiving the echo signal indicates that there is obstacle in this distance. After CPU has obtained the effective signal, CPU regulates the stall that is R-1 file to transmit corresponding current immediately and then detect whether there are more obstacles ahead. If the echo signal is not detected, it indicates that there is no obstacle in the corresponding distance between the R-file and R-1-file, CPU commands warning information. Receiving the echo signal indicates that the obstacle may be in any stalls from first stall to R-1-file. CPU controls the brake system to brake and keep the brake status, checking the speed until the speed dropped to k-file where there is no obstacle, then resume normal operation.

## 3 The Hardware Design

The operational amplifier, which has a high-gain, high CMRR (common-mode rejection ratio) paper size, amplifier will be used to expand the range of output current in this paper. It also could reduce the impact of noise and drift on the output current and linearity.

Infrared ranging circuit utilize the op-amp(operational amplifier) constant-current source circuit as its main component, which can change in external conditions in the other case, to provide a stable output current for load resistance, especially for the infrared light-emitting diode.

In order to output 500mA current, it must connect emitter-follower to the output of op-amp. A single transistor with emitter-follower is lack of drive current and the

Darlington transistor has a small adjustable range and bad flexibility. So we use compound transistor combined of two common transistors to output driving current. If a power transistor is used, it will require relatively large base current which drives the power transistor. But the op-amps generally work in the state of small current, which cannot drive the power transistor. If the op-amps are forced to drive the power transistor, it will cause high power consumption, high temperature-rise and affecting the circuit's output. So we use low power transistor drive high power transistor in order to meet the request of the design.

Slew rate is the maximum rate of rise of the output voltage per unit time. It is measured in  $V/\mu s$ . If a sharp step input voltage is applied to an op-amp, the output will not rise as quickly as the input because the internal capacitors require time to change to the output voltage level. SR is a measure of how quickly the output of an amp can change in response to a change of the input frequency. The SR depends on the voltage gain, but it is normally specified at unity gain.

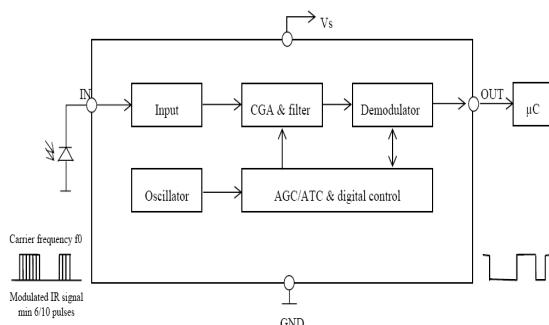
Slew rate can introduce a significant error if the rate of change of the input voltage is more than the SR of the op-amp. The rate of change of the input voltage rather than the change indicates how fast the input can rise.

If the rate of change of an input voltage is higher than the SR of the amplifier, the output will be highly distorted. SR is given by

$$SR = 2\pi f_i V_{om} \quad (1)$$

which gives the maximum frequency of the input voltage. From (1), it can be seen that the input frequency depends on the op-amp's SR. the higher the slew rate is, the greater the frequency of the input will be. So, in this paper, the SR of the op-amp reaches at least  $0.8V/\mu s$ . Thinking all these factors, we chose LM358.

In the negative feedback circuit, the output current changes are detected by sampling resistor which is in series with the load circuit. So as not to affect the performance of the constant current source, the stability of the sampling resistor is very important. However, the changes in temperature and effective will cause the value of resistance to change, and the sampling resistor should also have enough power, otherwise the performance of current source will be affected or burned.



**Fig. 1.** The diagram of the infrared-receiver demodulated output signal

Infrared-receiver receives the echo signal reflected by obstructions and completes the photoelectric conversion and has a function of PWM signal demodulation. The module receives 38 kHz continuous pulse signal. It can demodulate effectively and output a negative pulse as shown Fig.1. Receiver where J-V0, J-GND and J-IN respectively corresponds to power, ground, signal output has three pins as shown in Fig.1.

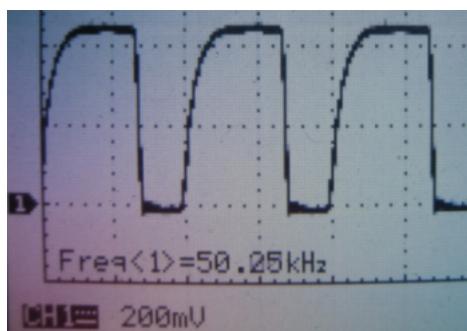
## 4 Experimental Results

To illustrate the importance of the SR of the op-amp, we make a systematic test for the infrared range with high-frequency current source. We select three values of input. The output waveforms are shown in Fig.2 and Fig.3.

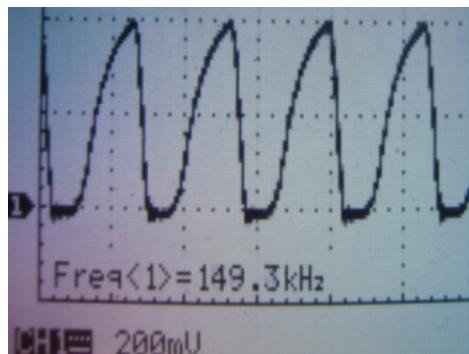
**Table 1.** Test Distance Data

Distance(m)	Frequency(kHz)	Peak current(mA)
20.8	150	250
21.22	150	250
20.9	150	250
21.1	150	250
21.3	150	250
55.8	50	500
56.1	50	500
55.3	50	500
55.5	50	500
56.0	50	500

From Fig.2 and Fig.3 we find the distortion of output waveform is more serious as the frequency increasing. This is due to the limit of op-amp's SR. The measurement data is shown in Table1. The two tables show that for the same measured object the greater the detected current is, the farther the distance is. During the experiment, there is no damage to the infrared diode, so it can be up to achieve the purpose of 500mA of output current.



**Fig. 2.** The output waveform of  $u_i=2v, f=50\text{kHz}$



**Fig. 3.** The output waveform of  $u_i=500\text{mv}$ ,  $f=150\text{kHz}$

## 5 Conclusions

In a traditional infrared ranging system, the avalanche transistor is commonly used as circuit driving the infrared diode. But it has not enough driving current and is unstable at signal with high frequency. And it is difficult to achieve long-distance infrared range. In this paper, we present an infrared ranging system based on high-frequency current source. The system uses C80501F MCU as the main controller chip, combined with the external interface circuit to achieve the functions, such as infrared transmitter, receiver and brake control. A large number of results show that when the input frequency is 150kHz, the circuit can continuously adjust the gear of output and the maximum current is up to 500mA. The system can be used in automotive infrared ranging system and has an extremely important practical value.

## References

1. Bruun, E.: High speed, current conveyor based voltage mode operational amplifier. *Electron Lett.* 28, 742–744 (1992)
2. Lytrivis, P., Tsogas, M., Thomaidis, G.: A Vehicular Filter Suitable for Co-operative Automotive Safety Applications. In: 2010 IEEE Intelligent Vehicles Symposium University of California, pp. 21–24 (2010)
3. Yi, X.: Design of a High Precision Program-controlled Constant-current Source. *Process Automation Instrumentation* 30, 63–65 (2009), doi:CNKI:SUN:ZDYB.0
4. Mueller, S., Ritter, H., Rohing, H.: Pre-crash application for multiple target situations. In: International Radar Symposium, pp. 1–4 (2006)
5. Jiang, H.: Research of a New Type of Constant Current Source with High-precision and High Temperature Stability. *Modern Electronics Technique* 77(008), 3–5 (2008), doi:CNKI:SUN:XDDJ.0.14-003
6. Zhao, D., Guo, R., Zhao, Y.: Design and Realization of the Digital Controlled DC Current Source Based on SCM. *Instrumentation Technology* 8, 58–60 (2008), doi:CNKI:SUN:YBJI.0.06-023

# Fault Diagnosis for Power Equipment Based on IoT

Yusheng Zhu, Xiaoqing Huang, Junyong Zhang, Jie Luo, and Jie He

College of Electrical and Information Engineering,  
Hunan University, 410082 Changsha, China  
royzhu2011@163.com

**Abstract.** The method of fault diagnosis for power equipment (PE) based on single source information has its uncertainty and inaccuracy, and the relation between symptoms and faults is complex and uncertain. So, the fault should be described by multiple and different characteristic information, and the idea of internet of things (IoT) is introduced into the fault diagnosis for PE. IoT can provide multi-characteristic information for fault diagnosis, including on-line monitoring information and patrol information, and more accurate and reliable diagnosis results can be obtained by handing and processing the information with the help of information fusion. In the paper, IoT and multi-source symptom information are introduced firstly. Then, the structure of information fusion is built. Finally, a simple architecture of fault diagnosis system for PE based on IoT is presented.

**Keywords:** IoT, multi-source information, information fusion, fault diagnosis, RFID, WSN, on-line monitoring, intelligent patrol, smart grid.

## 1 Introduction

Fault diagnosis for PE is very significant to the operation of the power system. Because of uncertainty of symptom information and complexity between symptoms and faults, the fault diagnosis method based on single-source information is neither accurate nor reliable [1-5]. What is worse, it may lead missed-diagnosis and misdiagnosis [1]. Thus, synthetic disposal and cooperative analysis for multi-characteristic signal of PE are needed.

In reference [2-4], the method of fault diagnosis for PE based on electrical tests is introduced, and multi-characteristic information can be acquired, but the tests need to be undertaken regularly and don't consider actual work situation of PE. So, it is possible to cause unnecessary shutting down and maintain PE behind time when some fault happens. With the development of sensor, wireless sensor network (WSN) and communication technology, etc, more and more sensor devices are introduced into on-line monitoring of PE, so diverse status information can be acquired timely and accurately under the condition of power on.

Similarly, Patrol information of PE is an effective basis of fault diagnosis, and it depends on human senses, such as sound, vibration, smell, color, to diagnosis PE. But this method adopts paper medium to record information, it may cause patrol

information missing and processed not timely. With the help of radio frequency identification (RFID) technology, patrol information can be transmitted expediently to patrol system and fault diagnosis system.

As a new generation information communication technology, IoT has attracted much attention of industry, academia and governments [6]. Because of its strong capacities of information perception and collection, IoT can acquire accurate status information of the objects and provide plentiful multi-source information related with faults for diagnosis by deploying multi-type sensors and utilizing RFID devices. So, multi-source symptom information can be guaranteed [7].

Information fusion is an information processing method for uncertain problems. It can lower the uncertainty, describe the objects more comprehensive, and get more precise and reliable diagnosis results by processing multi-source information than the method based on single-source information [1-5].

A new fault diagnosis method for PE based on IoT is proposed in this paper. This method makes full use of the redundant and complementary of on-line monitoring and patrol information, which are derived from IoT, and fuses the information with the help of some intelligent algorithms, which makes the diagnosis results more reliable, and the maintenance work will become more convenient.

## 2 IoT

IoT is a new burgeoning information communication technology, combined with a lot of technologies, such as WSN, RFID, communication and intelligent computing, etc.

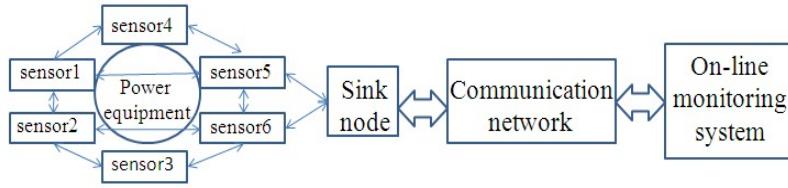
### 2.1 Conception of IoT

Generally, IoT is a thing-to-thing network on the basis of Internet that can connect every object through information sense devices, exchange information and communicate with each other, to identify, track, monitor and manage the object [8].

In power industry, IoT can realize intelligent application oriented smart grid, taking advantage of multi-type sensors and RFID devices, communication networks and combining information processing methods. IoT can greatly improve the capacities of information perception and exchange, and can be extensively applied in every link of power grid, including generator, transmission, transformation, distribution and consumption. So every part of smart grid is integrated seamlessly.

### 2.2 WSN

WSN is a self-organized network composed of a lot of sensor nodes, which have the capabilities of sensing, computing and wireless communication. WSN can not only acquire and transmit the monitoring information of target area through the cooperation of sensor nodes, but also can simply process and manage the collected information.



**Fig. 1.** The structure of on-line monitoring system based on WSN

In smart grid, WSN will have a good application prospect. In reference [11], the application of WSN in condition based maintenance, intelligent metering, intelligent home, fault location is discussed; the application of WSN in on-line monitoring for power transmission and transformation equipment is emphatically introduced in reference [8].

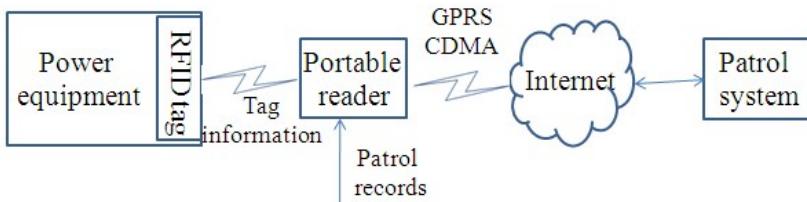
The on-line monitoring system based on WSN is shown in Fig. 1. Multiple status information of the equipment can be collected by deploying multi-type sensors, and sent to the sink node by means of multi-hop transmission, and finally access to the on-line monitoring system and data centre [9].

### 2.3 RFID

RFID is an automatic identification technology. It can automatically identify the signed object and acquire its corresponding information. A RFID based system is composed of electronic tag, reader and background management system.

In power system, RFID mainly is applied in intelligent patrol and asset management of PE [7-10]. As an important RFID device, portable readers can easily connect with Internet, which makes it convenient to apply in intelligent patrol.

The structure of intelligent patrol system based on RFID is shown in Fig. 2, when inspection personnel execute the patrol task, they use portable reader to identify the electronic tag of PE and acquire information of the tag, check the status item by item on the basis of corresponding resolved information and record the result of every checked item. If some term shows unusual, the abnormal information will be recorded and sent to the fault diagnosis system.



**Fig. 2.** The structure of intelligent patrol system based on RFID

### 3 Multi-source Information

Multi-source information is derived from many aspects. In this paper, multi-source symptom information for fault diagnosis provided by IoT mainly comes from two aspects: on-line monitoring information and patrol information.

#### 3.1 On-Line Monitoring Information

On-line monitoring is an important application of IoT in power grid. Because of its characteristics of real-time, continuity, and convenience (no need to shut down), power department is putting to use on-line monitoring devices instead of electrical tests to get status information of PE.

In on-line monitoring system, various, abundant characteristic information can be obtained. Take power transformer for example, characteristic information includes content of dissolved gas in oil, water content in oil, dielectric loss factor, partial discharge, winding deformation, etc [3]. If some data exceeds the set value and presents abnormal, we will choose it as a symptom. And finally symptom data will be sent to the fault diagnosis system.

#### 3.2 Patrol Information

Portable RFID reader can identify electronic tag automatically, record and save patrol information, and transmit the recorded information timely.

Patrol is an approach for fault detection of PE, and abnormal condition can be judged based on human senses. Similarly, some abnormal patrol information should be chosen as symptom. When diagnosing a power transformer, abnormal color, temperature and level of oil, and abnormal sound are usually selected as symptom information, referring some standards related with patrol of PE [3].

## 4 The Structure of Information Fusion for Fault Diagnosis

Information fusion is an information processing method, which analyses and comprehensively coordinates and optimizes the information and data from sensors to get the needed decisions and estimations based on some norms [1].

In fault diagnosis, information fusion can make full use of symptom information from multiple sources, to get the redundant and complementary of the information and describe comprehensively the fault of PE and improve the accuracy and reliability of diagnosis result.

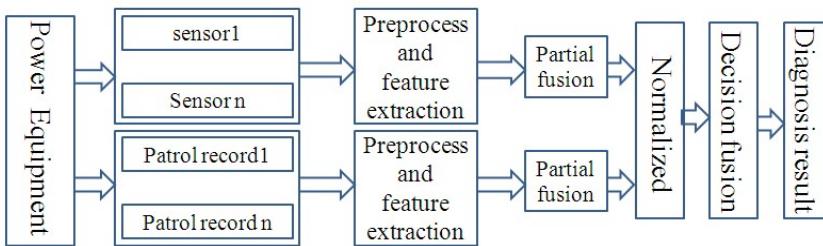
When using information fusion to diagnose PE, the first step is to establish the set of symptoms and faults. The fault set can be determined according to fault records of PE. Because of symptom information derived from two aspects, the symptom set is divided into two subspaces: on-line monitoring subspace and patrol subspace. In some conditions, the on-line monitoring data can be further divided. Take transformer for example, the on-line monitoring space can be divided into oil chromatogram subspace and other monitoring subspace.

The second, multi-source information should be preliminary processed and the feature will be extracted before fusion. For patrol information, it is the values of linguistic assessment, so it should be quantified. The method commonly adopted is Bipolar Scaling. The number 1, 2, 3, 4, 5, respectively stands for the change degree of PE's symptom parameter: no change, no significant change, change, significant change, very significant change [3].

For on-line monitoring information, it also needs to be transformed into multi-characteristic parameters, with the help of some mathematical methods, such as Fourier transform, wavelet analysis, etc [1]. So they can be conveniently processed.

Then, treating on-line monitoring symptom parameters and patrol symptom parameters respectively as the input of partial fusion, and using intelligent fusion algorithms, such as fuzzy theory, artificial neural network, rough set theory, etc, the respective failure probability of each subspace can be obtained.

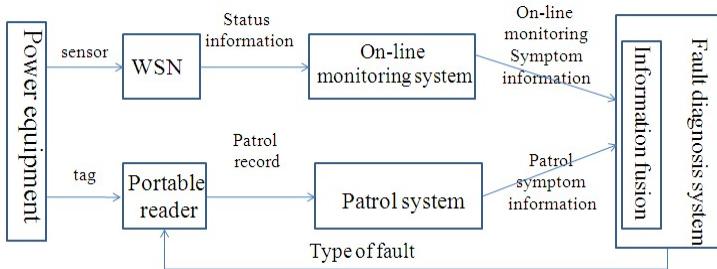
Finally, the diagnosis result will be integrated by processing the normalized data with the decision fusion methods, such as D-S evidence theory, Bayesian theory, etc. The structure of information fusion for fault diagnosis is shown in Fig. 3.



**Fig. 3.** The structure of information fusion for fault diagnosis

## 5 The Architecture of the Fault Diagnosis System for PE

The architecture of fault diagnosis system for PE is shown in Fig. 4. When inspection personnel check the status of PE, they use portable reader record the result of each checked item and send it to the intelligent patrol system by means of wireless communication, and the recorded data will be saved in the corresponding database. When some fault happens, the patrol symptom data will be transmitted to the fault diagnosis system. Simultaneously, on-line monitoring symptom information can also be transmitted to the fault diagnosis system. So, both patrol records and on-line monitoring information are processed by information fusion. Finally, the type of fault can be diagnosed and some decision will be made for maintenance, with the help of expert system, and all the information will be sent to portable reader by short message. So inspection personnel can easily know the type of fault and they can conveniently maintain PE.



**Fig. 4.** The architecture of fault diagnosis for PE

## 6 Conclusion

In this paper, a new fault diagnosis method for PE is proposed. The idea of IoT is introduced into fault diagnosis, and it can provide on-line monitoring information and patrol information for fault diagnosis. With the help of information fusion, all the information is processed, and more accurate diagnosis result can be obtained by fusing the results of partial fusion. Similarly, the maintenance group can easily get the type of fault, so, the faulted PE can be maintained in time.

**Acknowledgements.** Supported by the National High Technology Research and development Program of China (863 Program) (Grant No.2011AA05A120) and Hunan Province Research and Innovation Fund Project of Postgraduates (Grant No.CX2011B144).

## References

1. Zhu, D.Q., Liu, Y.N.: Information Fusion Method for Fault Diagnosis. *Control and Decision* 22(12), 1321–1328 (2007)
2. Shang, Y., Yan, C.J., Yan, Z., Cao, J.L.: Synthetic Insulation Fault Diagnosis Model of Oil-immersed Power Transformers Utilizing Information Fusion. In: Proceedings of CSEE, vol. 22(7), pp. 115–118 (2002)
3. Du, L., Yuan, L., Wang, Y.Y.: Power Transformer Fault Fusion Diagnosis Using FMADM Theory. *Journal of Chongqing University* 33(12), 1–7 (2010)
4. Liao, R.J., Liao, Y.X., Yang, L.J., Wang, Y.Y.: Study on Synthetic Diagnosis Method of Transformer Fault Using Multi-neural Network and Evidence Theory. In: Proceedings of CSEE, vol. 26(3), pp. 119–124 (2006)
5. He, J.J., Zhao, L.: Hydroelectric Generating Sets Fault Based on Information Fusion Technology. *J. Cent. South Univ. (Science and Technology)* 38(2), 333–338 (2007)
6. Zhu, H.B., Yang, L.X., Zhu, Q.: Survey on the Internet of Things. *Journal of Nanjing University of Posts and Telecommunications (Natural Science)* 31(1), 1–9 (2011)
7. Guo, C.X., Gao, Z.X., Zhang, J.J., Bi, J.Q.: IOT Based Transmission and Transformation Equipment Monitoring and Maintenance Assets Management. *Journal of Electric Power and Technology* 25(4), 36–41 (2010)

8. Wang, C.X., Yang, H., Wang, H.J., Zhang, J.Y.: Application of IoT in Condition Monitoring and Life Cycle Management for Power Transmission and Transformation Equipment. *Telecommunication for Electric Power System* 32(223), 116–122 (2011)
9. Lu, Z.J., Huang, R.H., Zhou, Z.Y.: Application Prospects of Internet of Things in Smart Grid. *Telecommunication for Electric Power System* 31(213), 50–52 (2010)
10. Su, Y.F., Xu, L.: Intelligent Optical Cable Patrol System Based on RFID. *Optical Communication Technology* 32(4), 35–37 (2007)
11. Wang, Y.G., Yin, X.G., You, D.H.: Application of Wireless Sensor Networks in Smart Grid. *Power System Technology* 34(5), 7–11 (2010)

# Fast QR Code Image Process and Detection

Qichao Chen, Yaowei Du, Risan Lin, and Yumin Tian

School of Computer Science and Technology,  
XiDian University, P.R. China 710071  
2009chenqc@163.com

**Abstract.** In this paper, we focused on how to improve the recognition rate of the QR Code in embedded systems of ordinary CMOS camera, and proposed an efficient pre-processing and detecting method for QR Code images with complex background or uneven illumination and an image binarization algorithm based on image blocks. Moreover, QR Code images which have geometric distortion or rotation can be fast corrected with our perspective transform matrix created by QR Code's finder patterns and alignment pattern. Experiments on WinCE embedded platform show that our image pre-processing and detecting methods can improve the recognition rate and accelerate the speed of the QR Code decoding.

**Keywords:** QR Code, Image Pre-processing, Perspective Transform, QR Code Location, QR Code Detection.

## 1 Introduction

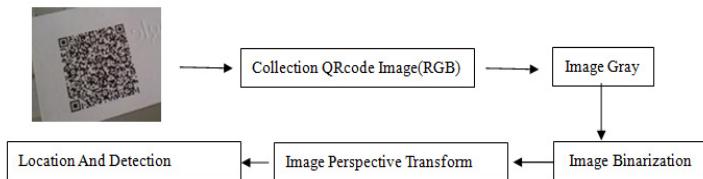
QR Code (abbreviated from Quick Response Code) is the trademark for a type of two dimensional code that designed for automotive industry. QR Code was invented by the Toyota subsidiary Denso-Wave in 1994. Recently, QR Code has been popular in Mobile Internet and the Internet of Things due to its fast readability and large storage capacity . Compared to the UPC barcode, QR Code obtains higher fault tolerance and the ability of error correction. Nowadays, QR Code has being increasingly widely used around the filed of e-commerce, electronic card and social network services.

So far, the researches on QR Code detecting and image pre-processing method are mostly based on rotated interpolation and Hough Transform. Jun Li adopting Hough transform based on edge detection and deformity correction as image pre-processing. Tingting Huang proposed a method that improved the Hough transform and noise filtering as image processing. Ming Sun proposed a method that improved the adaptive algorithm and hollowed contain to get code edge for Hough transform. However, the recognition rate of these methods is low. There are two defects : (1) rotation and interpolation can bring on additional image distortion. (2) Hough transform algorithm is an effective way to get the right edges of QR Code but is computationally expensive. In this paper, we propose an image pre-processing algorithm which is

based on perspective transform and can fast correct the QR Code location and detect the QR Code. The principal advantages of this method are that it has higher recognition rate and less image processing steps which result in less processing errors. The purpose of our algorithm is to solve the QR Code detection problem in its practical application.

## 2 Image Pre-processing

Image pre-processing is a critical step in the procedure of QR Code recognition. The results of image pre-processing have a vital impact on the QR Code recognition rate. The algorithm proposed for QR Code recognition process is based on the following flow diagram in Fig 1:



**Fig. 1.** Image pre-processing flow diagram

### 2.1 Converting to Gray-Level Image

The image captured by the embedded system in our platform is in RGB format. However, most information of the image in RGB color space is redundant to QR Code detection, as it needs more storage and is computational costly when compared with-level gray image. Thus the first step of image processing is to convert RGB color image into gray-level image. The classic converting algorithms are: (1) the Component Method, which selects a component from RGB's values as the gray value. (2) The Max-Component Method, which choose the maximum value of RGB values as the gray value (3) the Mean Method, which set the average value as the gray value. (4) Due to the fact that the human eyes is more sensitive to green, and less to blue, we use the following formula to calculate the gray-level image.

$$F(x,y) = 0.30R(x,y) + 0.59G(x,y)+0.11B(x,y) \quad (1)$$

### 2.2 Image Binarization

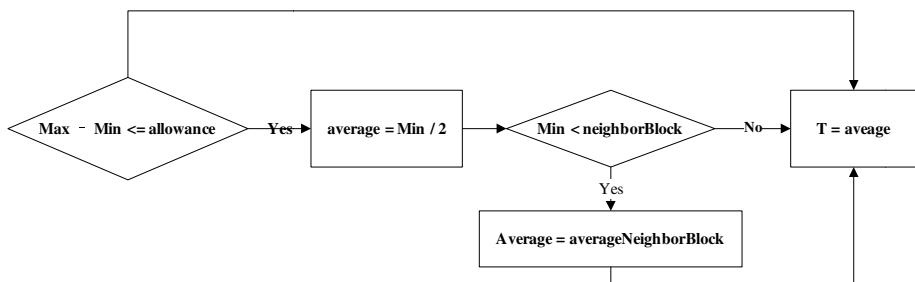
Binarization is a critical step in image pre-processing, the quality of image after binarization is also closely related to the recognition rate. Select an appropriate threshold is an important factor of binarization.

### 2.2.1 Image Binarization Theory

Due to the feature that QR Code image contains only black and white pixels, we need to use image binarization to extract the QR Code from background while protecting its structural integrity as possible as we can. The main idea is to define a threshold parameter  $T$ , then use  $T$  to divide the image pixels into two parts, the black and the white. Through thresholding, the binary image can be produced. The thresholding formula is as below:

$$F(x,y) = \begin{cases} 1 & , G(x,y) \leq T \\ 0 & , G(x,y) > T \end{cases} \quad (2)$$

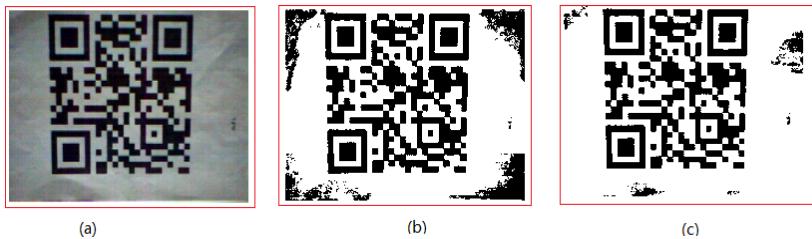
At present, according the different images and purposes, the thresholding methods can be mainly divided into three types: the Global Thresholding method, the Local Thresholding method and Dynamic Thresholding method. The references used OSTU or improved OSTU algorithm to binary the image. Under the complex circumstance of uneven illumination, Global Thresholding method is not ideal in performance. Therefore, according to the image-acquisition environment and the application platform, the method proposed in this paper obtains the binary image based on Local Thresholding, which can effectively retain all the features of QR Code under the uneven or complex environment. This method mainly consists of the following steps: firstly, the image is segmented into sub-blocks and each block contains 64 pixels with the size 8\*8. The purpose of dividing into blocks is to tackle the problem of uneven illumination; secondly, a threshold  $T_{i,j}$  is calculated in each block. As is shown in Fig. 2, our algorithm searches and finds the maximum, minimum and average gray value in each block, with which a reasonable threshold can be obtained.; finally, the binary image can be obtained by applying the threshold  $T_{i,j}$  to each block.



**Fig. 2.** Calculate the threshold in each block

### 2.2.2 Binary Image

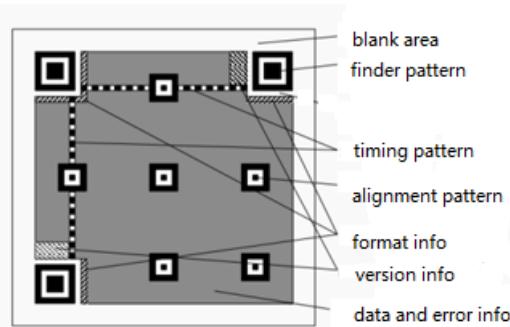
According to the formula (2) and the local threshold matrix  $T$ , we can convert the gray-level image into binary image. Since black block of QR Code represents value 1 and white 0, we set bit matrix that gray value less than threshold 1 and greater than threshold 0. Figure 3 as below shows the different effects of different threshold algorithms. It is evident that our method has a good effect to the source image.



**Fig. 3.** (a) Original Picture (b) global OSTU algorithm (c) Our method

### 3 QR Code Location

Figure 4 shows that QR Code contains the information of the finder pattern, timing pattern, alignment pattern, version information, error information and so on. There are many methods to locate and extract the QR Code from a camera picture. Tingting Huang proposed an algorithm based on the characteristics of QR Code, which detects the QR Code using different edge detection operators and finds the QR Code through Hough transform. Ming Sun put forward an algorithm that hollows out the pixels inside QR Code to get the edges, and finally uses the Hough transform to rectify the distorted image. Jinwei Wei look for position detection patterns before decoding by using opening and closing operator in morphological filtering. However, the recognition rate of the above methods is not high, as there are new errors brought in by the edge detection and Hough transform which also are computationally costly . Therefore, based on finder patterns and alignment pattern in the QR Code, an algorithm is proposed to create a perspective transform, with which we can effectively and efficiently extract the QR Code from the image.

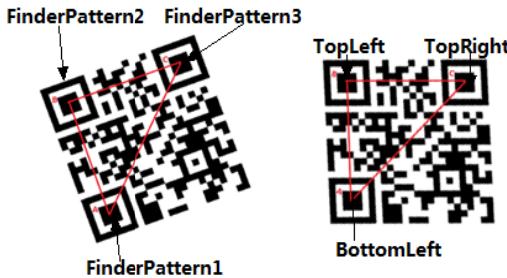


**Fig. 4.** QR Code information distribution

#### 3.1 QR Code Finder Pattern Location

The finder pattern is of ratio one-one-three-one-one in black and white, thus we can search in the binary image and locate the coordinates of the finder patterns which are

denoted as  $\text{FinderPattern1}(x,y)$ ,  $\text{FinderPattern2}(x, y)$ ,  $\text{FinderPattern3}(x, y)$ . Fig 5. (a) below shows the wrong positions of three finer patterns, after we find three finder patterns, we should transform the coordinates of the finder patterns to their right positions. According to the Pythagorean theorem, we can easily adjust the finder patterns to the scheme in Fig 5. (b)



**Fig. 5.** (a) the finder patterns that we found(b) the right position of three finder pattern

### 3.2 QR Code Alignment Pattern Location

When the version of QR Code is greater than 1, QR Code has more than one alignment pattern, and the only one we need to locate is the bottom-right alignment pattern which is denoted as  $\text{AlignmentPoint}(x, y)$ .

However, when the version is 1, QR Code does not have any alignment patterns, and the bottom-right point is just regarded as the alignmentPoint instead. We use the three finder patterns we find above and calculate a false appearance “finderPattern” and suppose it is the alignment Point.

## 4 QR Code Perspective Transform and Detection

The images of QR Code are always rotated or geometric distorted when we captured them. Therefore, before decoding it, we have to correct the QR Code image. In this paper, in order to create the perspective transform matrix, we just use the standard QR Code matrix sampling grid and the obtained patterns, including the three finder patterns and one alignment pattern.

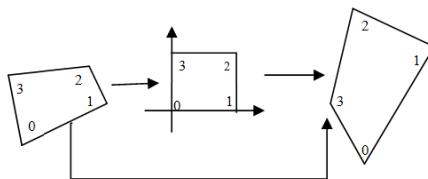
### 4.1 Perspective Transform

The general representation of the transform is illustrated as follows:

$$[x', y', w'] = [u, v, w] \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad (3)$$

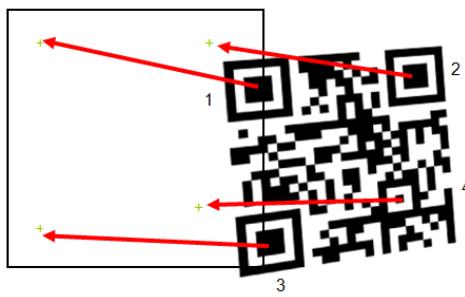
$$\text{where } \mathbf{x} = \frac{\mathbf{x}'}{w'}, \mathbf{y} = \frac{\mathbf{y}'}{w'}$$

We can establish the quadrilateral-to-square, square-to-quadrilateral and quadrilateral-to-quadrilateral perspective transform if we know the four source points and destination points. Figure 6 shows the scheme of the perspective transform.



**Fig. 6.** Perspective Transform

Based on QR Code three finder patterns and one alignment pattern, this paper proposed a method to correct the geometric distortion. Figure 7 as below shows the transformation process, and the procedure of the algorithm is as follows: firstly, according to the coordinates of three finder patterns and one alignment pattern, we establish the perspective transform formula and get the parametric values of transformation; secondly, according to the QR Code information and their corresponding locations, we can get the version information and dimension of QR Code and set up a null matrix based on the standard of this version; finally, map the points of the standard sampling grid of the QR Code to the original image through the perspective transform formula and get the value of module block, then set standard QR Code matrix value we have got.

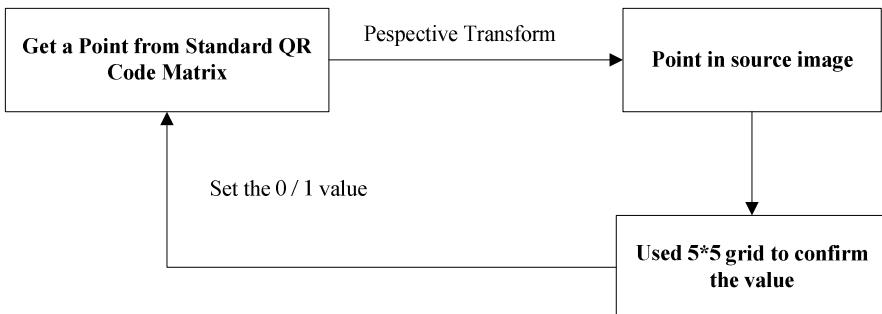


**Fig. 7.** Perspective Transform

## 4.2 QR Code Detection

The purpose of QR Code detection is to figure out the 0,1 values of the QR Code matrix. We proposed an algorithm that just uses perspective transform and standard QR Code matrix to accomplish the detection. The algorithm is as follows: firstly,

sampling grid the standard uses  $0.5 \times 0.5$  grid for each point and find the actual position in the source image through perspective transform. secondly, calculate the actual point value used  $5 \times 5$  sampling grid to confirmed the the value of QR Code. Finally, set the value in standard QR Code matrix. Figure 8 shows the detection process.



**Fig. 8.** QR Code detection process

## 5 Experimental Results

In order to test and verify the recognition rate and speed of our method, the image-acquisition environments of CMOS camera in WinCE platform are even illumination and uneven illumination. From version 1 to 10, ten QR Code images are chosen in each with the resolution of  $240 \times 180$ . In experiments, we use the GB method (International Standard) and our method which contains the image pre-processing to detect the QR Code separately. Table 1 shows experimental results of the recognition rate of each method.

**Table 1.** The experimental results

Environment	GB Method	Our Method
Uniform illumination	81%	89%
Uneven illumination	63%	84%

## 6 Conclusion

In this paper, we research on how to recognize QR Code efficiently and effectively in the image with geometry distortion. and the algorithms proposed include image binarization, QR Code location , QR Code image perspective transform and QR Code detection.

The main innovations in this paper are as follows: firstly, local block thresholding is proposed to improve the critical step in image preprocessing, and the algorithm can retain the QR Code features well and remove the complex background; secondly, in order to correct the rotated or geometric distorted image and increase the recognition rate of QR Code detection, the three finder patterns and one alignment pattern are used to accomplish the perspective transform; finally, using the result of perspective transform, we can detect the image and get the standard QR code data and finish the decode image pre-processing.

Experimental results show that the proposed method can improve the QR Code recognition rate and accelerate image pre-processing speed.

## References

1. ISO/IEC: Information technology-automatic identification and data capture techniques-QR code 2005 bar code symbol specification. ISO, Switzerland (2006)
2. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 2nd edn. PHEI (2010)
3. Zhang, S.: Image Project: Image Process. Tsinghua University Press (2007)
4. Gu, Y., Zhang, W.: QR Code Recognition Based on Image Processing, pp. 733–735. IEEE (2011)
5. Liu, Y., Yang, J., Liu, M.: Recognition of QR Code with Mobile Phonesm, pp. 204–205. IEEE (2008)
6. Zhou, J., Liu, Y., Kumar, A.: Research on Distortion Correction of QR code Images, pp. 417–418. IEEE (2012)
7. Ohbuchi, E., Hanaizumi, H., Hock, L.A.: Barcode Readers using the Camera Devic. In: Mobile Phones. IEEE (2004)
8. The national quality technology supervision bureau. The People's Republic state standards-fast response matrix Code (QR Code). GB/T18284. China standard press, Beijing (2000)
9. Li, J.: Research and Implementation of Technology of Image Restoration and Identification of QR Code, pp. 16–19. Soochow University (2010)
10. Huang, T.: The research on the technology of QR Code recognition, pp. 29–30. Central South University (2008)
11. Sun, M., Fu, L., Yang, Y.: Image Analysis Method for QR Code's Automatic Recognition. Journal of University of Electronic Science and Technology of China, 1019–1020 (2009)
12. Wei, J., Dai, S., Mu, A.: Rectification And Localization of QR Code Image Based on Methematical Morpholoy and Hough Transformation. Computer and Information Technology, 33–35 (2010)

# Towards Adaptable Workflow Management System: Shark Enhydra

Lazarus Obed Livingstone Banda, Zuping Zhang, and Jing Xia

School of Information Science and Technology, Central South University,  
Changsha, Hunan, China  
Chigoba2004@gmail.com, zpzhang@mail.csu.edu.cn,  
xiajing610@163.com

**Abstract.** Workflow activities may have due-dates to facilitate timely services and maximum use of resources like time. Due-dates (which are very critical in business) may pass before the intended participants can view their respective tasks because some environments lack necessary bandwidth and the administrator may not be aware of it. This paper presents a solution to this problem by designing and developing an agent-based plug-in that informs workflow administrators about connectivity status to a workflow server. The plug-in comprises three agents which will interface with system through Common Object Request Broker. Agents are employed in the design and implementation of the plug-in to capitalize on the current design in that Shark Enhydra is employing the same agent-based technology. The three collaborating agents (together forming a sub-system) shall persistently be communicating with each other through remote procedure calls (RPC) only if there is network connection and communicate with the workflow system through CORBA architecture.

**Keywords:** Plugin, network connectivity status, distributed system, deadline, static agent, mobile.

## 1 Introduction

Workflow is an act of automating business process in part or in total through which documents, information or tasks are being routed from one participant to another for action, based on a set of prescribed rules [1]. Among other things, software development is primarily concerned with striking equilibrium between business rules and user requirements. End users may be interested more in their requirements to an extent of almost overshadowing the former. In many cases, such balancing is only between what the end-user wants to do with the system and what the system functional requirements are supposed to be. Sometimes, the actual working conditions in which such systems will be deployed have very unsupportive and unfavorable environment. For example, the software may be able to do what a user wants in one IT environment while the same system may fail in a different environment. Much as it is non-starter to design a system that can work in all environments, but our concern is that it should at least work under general average conditions worldwide.

Besides other factors, we need to consider time management flexibility, effective communication, user issues, and integration across different departments or organizations (e-Commerce) [1] in relation to real life situation in different operating environments. This paper lobbies for a balanced approach between business rules and user requirements when developing Business Workflow Management Systems (BWfMS) and taking a step further in that we advocate that such systems should deliver realistic system functionality requirements and drives optimal performance vis-à-vis prevalent IT environmental conditions. We refer to such a balance as the “Adaptable Workflow”.

In underdeveloped countries, business environment suffers chronic bandwidth and connectivity. For example, the sub-Saharan region has too low connectivity compared to the city of Johannesburg due to prohibitively exponential unaffordable high costs of internet services and substandard infrastructure. In 2007, 16 countries in Africa had just one international Internet connection with a capacity of 10 Mbps or lower, while South Africa alone had over 800 Mbps [2]. This is a vivid bottleneck to BWfMS. It clearly explains why, a large section of Africa’s network traffic goes through expensive satellite links thereby making internet accessibility unaffordable for most of the population. Therefore, software developers should consider all factors that might lead to bottlenecks for a consumer. This makes us ask a multi-million question “how could we tweak a Workflow Management System to reflect realities of the environment in which it is deployed, to align it with the available minimal network strength such that due dates do not result into penalizing innocent workflow participants?” This point brings us to issues about maximizing abilities and adaptability of systems [3]. The user’s local environment should receive appropriate attention when developing a workflow engine, and as such, components should not overburden the bandwidth, to accelerate business.

## 2 Related Works

Xinhua Jiang and Lina Zhang presented an interoperation model among distributed workflow engines based on Shark and Asynchronous Web Services to solve workflow engines inter-operation to achieve information interaction among distributed engines [4]. Moreover, Daniel Nurmi et al. worked on distributed computing environment [5]. Their paper adds an existing workflow scheduler by introducing methods that make accurate predictions of both the performance of the application on specific hardware, and the amount of time individual workflow tasks will spend waiting in batch queues. Of similar interest is a paper by Gurmeet Singh et al [6]. They noted that the typical wait time experienced by the tasks in the resource job queue is often much more than their runtime, culminating into a workflow completion time greater than what could be achieved on a dedicated system. The authors ascertain that these workflows have a high degree of parallelism and a large number of tasks can execute concurrently, just as in a situation where an organization shares the same network with other departments, this could reduce upload and download speeds. In order to cope, the resource management systems impose limitations on the number of tasks that a user can submit at a time thereby throttling the execution of the workflow. Is it a proven fact that the overall quantity of resources in a workflow process is limited by the cost of employment; meanwhile, the assignment of resources directly decides the average

waiting and dealing time of each activity in a workflow. Certainly, optimization of resource allocation could minimize the total flow time of a workflow. Tie-Nan Deng, who designed a mathematical model of resources optimization with minimal cost time, addressed some optimization rules for minimizing the average responding time of a workflow [7]. The work contributed much to improvement of waiting time. However, that works well for resources that we can easily manipulate. In case of network connectivity (as a resource) in the third world, this optimization would be joke of the day, as it shall have to take decades to register such optimal levels of network connectivity. Network reliability and bandwidth in poor countries as well as in the countryside of many developing countries still leaves a lot to be desired. Sometimes it cannot support some basic workflow operations. This is a serious problem and therefore we need to deal with it. This justifies the need for other measures for dealing with network connectivity problem in an organization. This is where our paper comes into play.

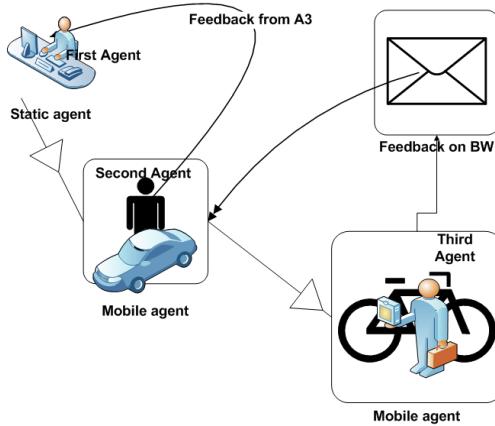
### **3 Our Approach and Solution**

As far as the current settings of the Shark Enhydra workflow enactment service are concerned, the deadline management is “static”. This means that once a workflow owner configures the deadline parameter, the task will expire as prescribed in the parameter setting, irrespective of the user’s network environment. It will not dynamically respond to the prevalent network connectivity status. Whether connectivity is available or not, time-out counter will start ticking even if the intended participants cannot log onto the system and access their respective tasks. This is a serious problem especially if the workflow is deployed in environments where network connectivity challenges are the order of the day. The objective of our research is to design an efficient Workflow Intelligence Module pluggable to a workflow enactment engine that will detect connectivity status (whether the network has problems or is in good working condition) to help the administrator or workflow owner make informed decisions about what action to take depending on the network status. The personnel operating the workflow at that point in time will have a range of options as to what action to take depending on the reported network status. Through this module design, we do not suggest any action to the administrator because solutions to any reported problems might differ from one situation to another. However, among many workable solutions, the administrator may opt to engage the services of a network systems administrator, request the workflow owner to redefine the tasks’ due-dates (a bit time consuming), suggest to the workflow owner to drop the workflow from the system (undesirable and uneconomical, sometimes), ignore any notifications from the plug-in, etc.

### **4 How the Module Operates**

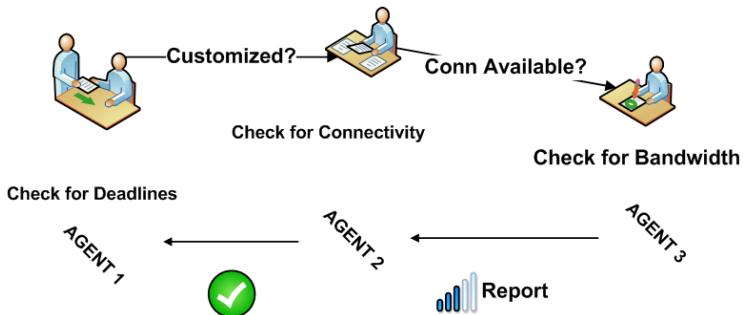
The module shall carry all its operations in the background such that the workflow administrator or the participants will not be aware of any operations about it except that it shall present the network status results to the local machine display. In this case, we shall have no menu or user interface to this module. The plug-in will be invoked

implicitly. As an agent-based plug-in subsystem, one agent will always be checking whether the workflow loaded into the workflow management system has customized due-dates for the workflow tasks (as might have been done at process definition stage).. If so, the plug-in will continue engaging its subsequent agents to perform their respective functions. The plug-in checks network connection to the local machine. The module shall comprise three agents. The first will be a static and the other two will be mobile as depicted in figure 1.



**Fig. 1.** Types of Agents in Our Plug-in

The first agent (the only static one) shall be charged with the task of checking if the deadline parameters of the loaded workflow tasks evaluates to a particular integer above zero. If this agent detects customization of deadline parameters, then it shall invoke the second agent. Refer to figure 2 for detailed roles of each agent in the plug-in.

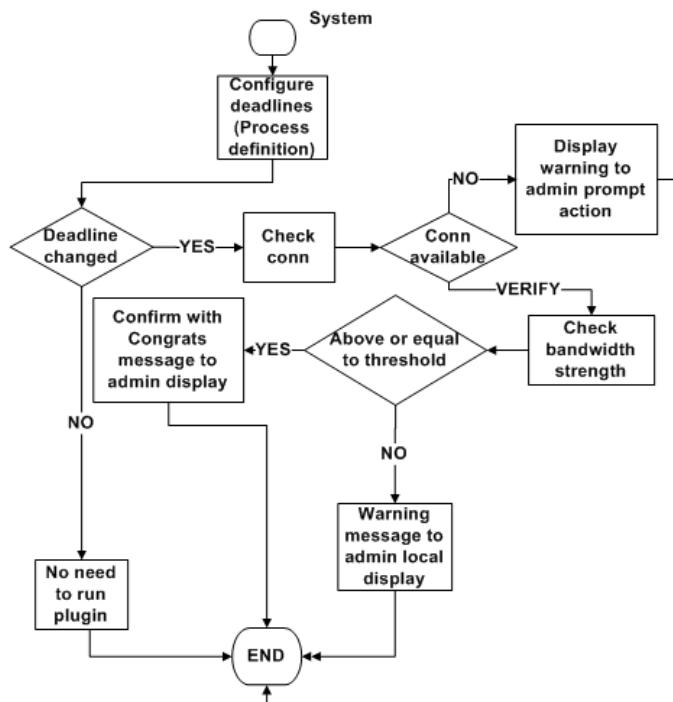


**Fig. 2.** Figure Relationships and Communication among Agents

## 5 The General Plug-In Design

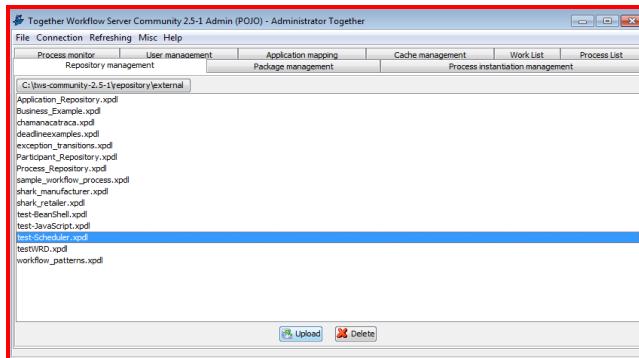
The first agent shall be waiting for feedback from the second. If there is no connection, the first agent waits in vain until time up. Then it composes a message to display that there is no connection. If there is no connection to the server, this agent

will suspend the timer on the local machine until a second agent reports connection restoration. The workflow owner or administrator will be notified immediately about the suspension of the countdown timer. The second agent is responsible for finding out if the local machine is connected to the server. In case there is connection detected, the second agent delegates (to the third) the responsibility of checking on the bandwidth and awaits feedback from the last agent. The third agent checks whether the available bandwidth is less than, equal to, or above a set threshold. In either case, the agent will inform the second agent, which will in turn inform the first agent to create appropriate messages for workflow administrator about the bandwidth status. Agents 2 and 3 will not give the administrator feedback directly. This is a deliberate design to avoid a scenario whereby network connection may be lost after the second agent has already verified it. (If the connection is lost at the time the second or third agent is working, then the first agent will not receive any information from the last two since message-passing channels may have been disconnected). So in such a case where there is no feedback until time out from subsequent agents to the first, then the first one assumes and reports that there's no connection. Since solely the first agent manages the responsibility of composing messages for the administrator, there will always be feedback to the administrator as this is a static agent on the local machine. It will be up to the administrator to take any action or ignore the message. This will enable the respective workflow participants to take action without unnecessarily locking them out of the system due to untimely deadlines. Refer to figure 3 showing the module design.



**Fig. 3.** Design of Workflow Intelligence Module Plug-in

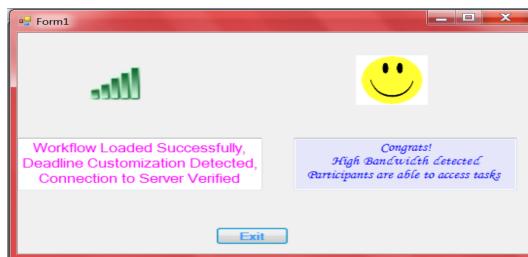
In figure 3, Conn means connection. The plug-in is triggered if and only if the deadline parameter is customized (if the default deadline settings are adopted, the plug-in remains inactive, in which case it implies that it does not matter whether the task will take so long to be accomplished by the intended participant(s)). If the deadline parameter is default, the screen looks just as in figure 4, with no feedback to the administrator.



**Fig. 4.** Results of loading workflow with default parameter

In any case, whether there is limited or no connectivity, or optimum bandwidth, WIM notifies the workflow administrator about the condition. We wish to emphasize that the module does not compel the administrator to take action of any type; neither does it give any suggestions. The administrator, may inform the network systems administrator, ignore the messages from the plug-in, suggest workflow process re-definition, or advise the workflow owners to drop the workflow.

The following screenshots (figures 5 through figure 7) show samples of the feedback messages from the activated plug-in. We will deploy our plug-in in open distributed systems architecture through CORBA. The design capitalizes on the advantages and strengths of multi agent systems like perseverance and persistence as there shall be need to check on the environment for connectivity continuously and persistently.



**Fig. 5.** High Bandwidth Feedback Sample Screenshot

If the plug-in reports high bandwidth connectivity, the administrator has no worries about the participants' access to the workflow tasks, though there is a

possibility that network connection may misbehave thereafter, after the confirmation message. At least our worry is letting the administrator about the current situation.

In a situation of low bandwidth (figure 6), the administrator may have to take measures that may help the participants have enough time to work on their online tasks. This may include recommending a process re-definition to redefine the deadlines. Ignoring the caution may be a possibility, hoping all will work so soon.

If the network connection is not available, the workflow administrator may wish to inform a network administrator to check on the connectivity status. In extreme situations, the administrator may advise on whether to drop the workflow or resort to manual workflow processing.



**Fig. 6.** Low Bandwidth Feedback Sample Screenshot



**Fig. 7.** No Network Connection

## 6 Conclusion and Future Works

Presented in this paper is a pluggable workflow intelligence module that can alert the workflow systems administrator about the state of network connectivity in order to respond appropriately where necessary. The Module does not oblige the administrator to take any action but helps the same to come up with informed decisions about the workflow to reflect realities of the IT environment in which the participant is operating. In this work, we have mentioned that the plug-in will have one of its agents notifying the workflow management systems administrator about the status of the network connection. In some instance, it will have to compare bandwidth against a threshold. Apparently, for experimental reasons, we chose our threshold arbitrarily,

but we would like to come up with authentic statistics and research as to what could be a threshold globally for general optimal bandwidth below which workflow processing may not be possible. We would like to come up with a mechanism that will make the plug-in determine and differentiate poor connectivity from high bandwidth. Meanwhile, we would also like to come up with a small project to enable runtime customization of the workflow due dates at execution time instead of always redefining the process. We deem this project a viable one in the interest of time.

**Acknowledgments.** We hereby thank Jane, Victory, and Peace for the encouragement and moral support during the whole period were preparing for this paper. This work is supported by the National Science Foundation of China (NSFC), Grant 60970095.

## References

1. Hollingsworth, D.: Workflow Management Coalition: The Workflow Reference Model Document Number TC00-1003 Document Status - Issue 1.1 (January 19, 1995)
2. Wikipedia. Internet in Africa (January 26, 2011)
3. Muller, R., Greiner, U., Rahm, E.: AGENT WORK: a workflow system supporting rule-based workflow adaptation. Data & Knowledge Engineering 51(2) (November 2004)
4. Jiang, X., Zhang, L.: Inter-operation of distributed workflow engine on asynchronous web services, Hohhot, China (2007)
5. Nurmi, D., Mandal, A., Brevik, J., Koelbel, C., Wolski, R., Kennedy, K.: Evaluation of a workflow scheduler using integrated performance modelling and batch queue wait time prediction (2006)
6. Singh, G., Su, M.-H., Vahi, K., Deelman, E., Beriman, B., Good, J., Katz, D.S., Mehta, G.: Workflow task clustering for best effort systems with Pegasus, USA (February 2008)
7. Deng, T.-N., Yi, Y., Chang, H.-Y., Xiao, Z.-J., Inoue, A.: Model and intelligent algorithm for workflow resource optimization to minimize total flow time, Dalian, August 13-16. IEEE (2006)

# Quantized Communication of Multi-agent Systems under Switching Topology<sup>\*</sup>

Qian Ye, Xuyang Lou, and Baotong Cui

Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education),  
Jiangnan University, Wuxi 214122, China  
yeqian6 @163.com

**Abstract.** In this paper, we discuss the problem of state agreement of multi-agent systems with quantized communication. Switching weighted topology is taken into account and we establish conditions for both uniform and logarithmic quantization. Convergence guarantees are provided when the graph for any arbitrary switching signal is a tree sufficiently often for the logarithmic quantizer, using algebraic graph theory. The results are illustrated through a numerical simulation.

**Keywords:** Multi-agent, quantized communication, switching topology, algebraic graph theory.

## 1 Introduction

The interaction between information flow and system dynamics has gained increasing attention recently. It is further recognized that information and communication constraints may have a considerable impact on the performance of a control system. The study of these topics forms new active areas of research such as *quantized control systems*, *networked control system* and *multi-agent systems*. In a dynamical system, the communication topology which determines what information is available for which component at a given time instant, is an important aspect of information flow. This topic on multi-agent system in this paper is motivated by its broad applications in many areas including cooperative control of unmanned air vehicles, formation control [1]–[3], flocking [4,5], sensor networks [6,7], attitude alignment of clusters of satellites, and congestion control in communication networks [8].

As is well known, communication constraints play a major role in consensus and related problems of distributed computation and control. Recently, the constraint of quantization, that is of communication restricted to a discrete set of symbols, has received significant attention [9, 10]. A number of works have been paid attentions to consensus or stability of distributed multi-agent networks under quantized communication, mostly in discrete-time systems, namely uniform randomized

---

\* This work is partially supported by National Natural Science Foundation of China (No.61174021, No.61104155), the 111 Project (B12018), and the Jiangsu Provincial Program for Postgraduate Scientific Innovative Research of Jiangnan University (No. CXZZ11\_0463).

quantizers [11], uniform deterministic quantizers [9, 12], logarithmic quantizers [13-15], and adaptive quantizers [16].

In this paper, we analyze the quantized average consensus of multi-agent systems using a continuous-time model. The switching weighted communication topology is considered with uniform and logarithmic quantizers and a connection between algebraic connectivity of the network and the performance of reaching an agreement is established. Due to the constraint of uniform quantization, we cannot obtain exact consensus, but we can obtain the convergence region of consensus errors.

## 2 Preliminaries

Let  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$  be a weighted digraph of order  $n$  with the set of vertices  $\mathcal{V} = \{v_1, \dots, v_n\}$ , set of edges  $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ , and a weighted adjacency matrix  $\mathcal{A} = [a_{ij}]$  with nonnegative adjacency elements  $a_{ij}$ . The vertex indexes belong to a finite index set  $\mathcal{I} = \{1, 2, \dots, n\}$ . An edge of  $\mathcal{G}$  is denoted by  $e_{ij} = (v_i, v_j)$ . The adjacency elements associated with the edges of the graph are positive, i.e.,  $e_{ij} \in \mathcal{E} \Leftrightarrow a_{ij} > 0$ . Moreover, we assume  $a_{ii} = 0$  for all  $i \in \mathcal{I}$ . The set of *neighbors* of node  $v_i$  is denoted by  $\mathcal{N}_i = \{v_j \in \mathcal{V} : (v_i, v_j) \in \mathcal{E}\}$ . A *path* of length  $p$  from  $v_i$  to  $v_j$  is a sequence of  $p+1$  distinct vertices starting with  $v_i$  and ending with  $v_j$  such that consecutive vertices are adjacent. For  $v_i = v_j$ , this path is a *cycle*. If there is a path between any two vertices of  $\mathcal{G}$ , then  $\mathcal{G}$  is *connected*. A connected graph is a *tree* if it contains no cycles. The *degree*  $d_i$  of vertex  $v_i$  is given by  $d_i = \sum_j a_{ij}$ . Let  $\Delta = \text{diag}(d_1, \dots, d_N)$ . The Laplacian of  $\mathcal{G}$  is the symmetric positive semidefinite matrix  $L = \Delta - \mathcal{A}$ . For a connected graph,  $L$  has a single zero eigenvalue with the corresponding eigenvector  $\mathbf{1} = [1, \dots, 1]^T$ . Let  $x_i \in \mathbb{R}$  denote the value of the node  $v_i$ . We refer to  $\mathcal{G}_x = (\mathcal{G}, x)$  with  $x = [x_1, \dots, x_n]^T$  as a network (or algebraic graph) with the value  $x \in \mathbb{R}^n$  and the topology (or information flow)  $\mathcal{G}$ . We say both the vertices  $v_i$  and  $v_j$  agree in a network if and only if  $x_i = x_j$ . We say the vertices of a network have reached a consensus if and only if  $x_i = x_j$  for all  $i, j \in \mathcal{I}$ .

Suppose the dynamics of each node follows

$$\dot{x}_i = f(x_i, u_i), i \in \mathcal{I}. \quad (1)$$

A dynamic graph (or dynamic network) is a dynamical system with a state  $(\mathcal{G}, x)$  in which the value  $x$  evolves according to the network dynamics  $\dot{x} = F(x, u)$ . Here,  $F(x, u)$  is the column-wise concatenation of the elements  $f(x_i, u_i)$  for  $i \in \mathcal{I}$ . In a dynamic network with switching topology, the information flow  $\mathcal{G}$  is a discrete-state of the system that changes in time. We say that, for all initial conditions, solutions to (1) satisfy the *average consensus*, if

$$\lim_{t \rightarrow \infty} x(t) = x_{\text{ave}}(0)\mathbf{1}, \quad (2)$$

where we let  $x_{\text{ave}}(t) = \mathbf{1}^T x(t) / n$ . Note that  $\mathbf{1}^T L = 0$  when  $\mathcal{G}$  is a *tree*. Thus,  $\alpha = x_{\text{ave}}(t)$  is an invariant quantity. The invariance of  $x_{\text{ave}}(t)$  allows decomposition of  $x(t)$  according to  $x(t) = \alpha \mathbf{1} + \bar{x}$ , where  $\bar{x} = [\bar{x}_1, \dots, \bar{x}_n]^T \in \mathbb{R}^n$  satisfies  $\mathbf{1}^T \bar{x} = 0$ .

### 3 Main Results

#### 3.1 Model

Consider  $N$  agents. Let  $q_i \in \mathbb{R}^2$  denote the position of agent  $i$ . Let  $x_i, y_i$  denote the coordinates of agent  $i$  in the  $x$  and  $y$  directions, respectively. Let  $z = [z_1^T, \dots, z_N^T]^T$  denote the vector of all agents' positions. We assume that agents' motion obeys the single integrator model:

$$\dot{z}_i = u_i, \quad i \in V = \{1, 2, \dots, N\}, \quad (3)$$

where  $u_i$  denotes the control input for each agent. We assume that each agent has limited information on the states and goals of the other group members. In particular, each agent is assigned a neighbor set  $\mathcal{N}_i \subset V$ , which is given by the agents with whom it can communicate.

Consider system (1) in the  $x$ -direction and let  $x = [x_1^T, \dots, x_N^T]^T$ . Without loss of generality, we omit the notation regarding the  $x$ -direction from the control input. We then have  $\dot{x}_i = u_i$ . We consider the following agreement control laws:

$$u_i = -c \sum_{j \in \mathcal{N}_i} a_{ij} (x_i - x_j) \quad (4)$$

where  $c > 0$  represents the coupling strength. Given protocol (4), the state of a network of continuous-time integrator agents evolves according to the following linear system:

$$\dot{x}(t) = -cLx(t) \quad (5)$$

where  $L$  is defined by

$$l_{ij} = \begin{cases} \sum_{k=1, k \neq i}^n a_{ik}, & j = i \\ -a_{ij}, & j \neq i \end{cases} \quad (6)$$

#### 3.2 Switching Communication Topology

In a network with *switching topology*, convergence analysis of protocol (4) is equivalent to stability analysis for a *hybrid system*:

$$\dot{x}(t) = -cL_{\sigma(t)}x(t) \quad (7)$$

where  $L_{\sigma(t)} = \mathcal{L}(\mathcal{G}_{\sigma(t)})$  is the Laplacian of graph  $\mathcal{G}_{\sigma(t)}$  that belongs to a set  $\Gamma$ . The set  $\Gamma$  is a finite collection of digraphs of order  $n$  with an index set  $\mathcal{I}_\Gamma \subset \mathbb{Z}$ . The map  $\sigma(t) : \mathbb{R} \rightarrow \mathcal{I}_\Gamma$  is a *switching signal* that determines the network topology.

Let  $\bar{x} = x(t) - x_{\text{ave}}(t)\mathbf{1} = x(t) - \alpha\mathbf{1}$  and note that  $\mathbf{1}^T L_s = 0$  when  $\mathcal{G}_s$  is a *tree* for any arbitrary switching signal  $s = \sigma(t) \in \mathcal{I}_\Gamma$ . Thus,

$$\dot{\bar{x}}(t) = -cL_{\sigma(t)}\bar{x}(t) \quad (8)$$

Here, we refer to  $\bar{x}$  as the (group) disagreement vector. The vector  $\bar{x}$  is orthogonal to  $\mathbf{1}$ .

Our goal is to construct a quantized communication strategy such that agents converge to the common value  $\alpha\mathbf{1}$  in the state space under quantized relative position information of their neighbors.

### 3.3 Quantized Control

In this paper, we consider two types of quantized sensors: uniform and logarithmic quantizer. They are given as:

(Q1) the uniform quantizer,  $q_u : \mathbb{R} \rightarrow \mathbb{R}$ ,  $\|q_u(a) - a\| \leq \delta_u, \forall a \in \mathbb{R}$ .

(Q2) the logarithmic quantizer,  $q_l : \mathbb{R} \rightarrow \mathbb{R}$ ,  $\|q_l(a) - a\| \leq \delta_l \|a\|, \forall a \in \mathbb{R}$ .

In the previous equations,  $\delta_u, \delta_l$  are positive scalar gains. We shall use the notation  $q(\cdot)$  for the quantizer when it is not specified if it is a uniform or a logarithmic quantizer. For a vector  $v = [v_1, \dots, v_d] \subset \mathbb{R}^d$  of size  $d$ , the following bounds are easily shown to hold:

(Q3) In the uniform quantizer case,  $\|q_u(v) - v\| \leq \delta_u \sqrt{d}$ .

(Q4) In the logarithmic quantizer case,  $\|q_l(v) - v\| \leq \delta_l \|v\|$ .

### 3.4 Quantized Agreement

In the case of quantized information we have

$$\dot{\bar{x}}(t) = -cL_{\sigma(t)}\bar{x}(t) \quad (9)$$

where  $q(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$  is the quantizing function. If this function satisfies  $q(-a) = -q(a)$  for all  $a \in \mathbb{R}$ , which is the case for both types of quantizers.

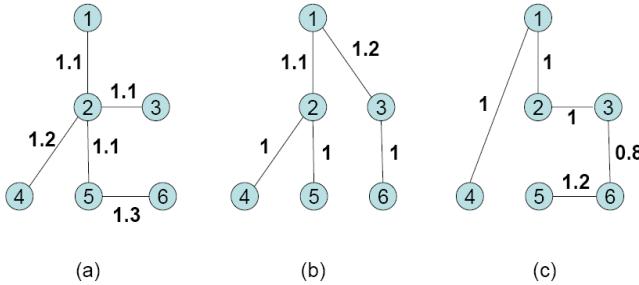
**Theorem 2:** Suppose that for any arbitrary switching signal  $s = \sigma(t) \in \mathcal{I}_\Gamma$  the switching communication graph  $G_s$  is a tree. Then the closed loop system (9) has the following convergence properties:

(i) In the case of a uniform quantizer, the system converges to a ball of radius  $\|L_s\| \delta_u \sqrt{m} / \lambda^*$  centered in the desired equilibrium point  $\bar{x} = 0$  in finite time, where  $m = |\mathcal{E}|$  is the number of edges.

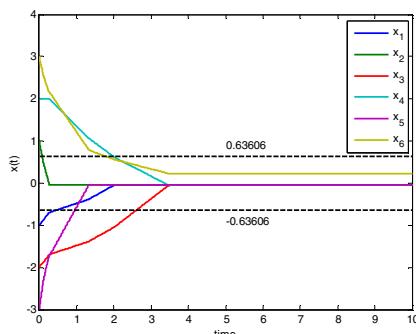
(ii) In the case of a logarithmic quantizer, the system is exponentially stabilized to an agreement point  $\bar{x} = 0$ , provided that the gain of the quantizer  $\delta_l$  satisfies  $\delta_l < \lambda^*/\|L_s\|$ .

## 4 An Example

Consider the multiagent system with weighted switching topologies  $\mathcal{G}_s$  ( $s = 1, 2, 3$ ) shown in Figure 1. As can be seen, the number of edges in all graphs is  $m = 5$ . We choose  $\delta_u = 0.05$  in the simulation. It is easy to verify that  $\max_{s=1,2,3}\{\|L_s\|\delta_u\sqrt{m}/\lambda^*\} = 0.63606$ . Therefore, according to Theorem 2, the system should converge to a ball of radius 0.63606 in the case of using uniform quantizer. To see the simulation easily, we set the initial states as  $x(0) = [-1, 1, -2, 2, -3, 3]^T$  satisfying  $x_{ave}(0) = 0$ . The state trajectories of the system with uniform quantization are shown in Figure 2. We find that the visual results illustrate the theoretical analysis.



**Fig. 1.** Three weighted graphs for switching cases



**Fig. 2.** State evolutions of uniform quantized dynamics on the switching graph in Figure 1

## 5 Conclusion

In this paper, quantized consensus protocols have been proposed for multi-agent systems with switching weighted topologies. Sufficient conditions for the convergence of the systems are obtained. Simulations have shown the effectiveness of the proposed consensus protocols.

## References

1. Fax, A., Murray, R.M.: Information flow and cooperative control of vehicle formations. *IEEE Trans. Automat. Contr.* 49, 1465–1476 (2004)
2. Olfati-Saber, R., Murray, R.M.: Distributed cooperative control of multiple vehicle formations using structural potential functions. Presented at the 15th IFAC World Congr., Barcelona, Spain (June 2002)
3. Vidal, R., Shakernia, O., Sastry, S.: Formation control of nonholonomic mobile robots omnidirectional visual servoing and motion segmentation. In: Proc. IEEE Conf. Robotics and Automation, pp. 584–589 (2003)
4. Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, O., Shochet, I.: Novel type of phase transition in a system of self-derived particles. *Phys. Rev. Lett.* 75(6), 1226–1229 (1995)
5. Toner, J., Tu, Y.: Flocks, herds, and schools: a quantitative theory of flocking. *Phys. Rev. E* 58(4), 4828–4858 (1998)
6. Cortés, J., Martinez, S., Karatas, T., Bullo, F.: Coverage control for mobile sensing networks. *IEEE Trans. Robot. Autom.* 20(2), 243–255 (2004)
7. Cortés, J., Bullo, F.: Coordination and geometric optimization via distributed dynamical systems. *SIAM J. Control Optim.* 44(5), 1543–1574 (2006)
8. Paganini, F., Doyle, J., Low, S.: Scalable laws for stable network congestion control. Presented at the Int. Conf. Decision and Control, Orlando, FL (December 2001)
9. Frasca, P., Carli, R., Fagnani, F., Zampieri, S.: Average consensus on networks with quantized communication. *International Journal of Robust and Non-Linear Control* 19(16), 1787–1816 (2009)
10. Kashyap, A., Basar, T., Srikant, R.: Quantized consensus. *Automatica* 43(7), 1192–1203 (2007)
11. Aysal, T.C., Coates, M.J., Rabbat, M.G.: Distributed average consensus with dithered quantization. *IEEE Transactions on Signal Processing* 56(10), 4905–4918 (2008)
12. Nedic, A., Olshevsky, A., Ozdaglar, A., Tsitsiklis, J.N.: On distributed averaging algorithms and quantization effects. *IEEE Trans. Automat. Contr.* 54(11), 2506–2517 (2009)
13. Carli, R., Bullo, F., Zampieri, S.: Quantized average consensus via dynamic coding/decoding schemes. *International Journal of Robust and Nonlinear Control* 20(2), 156–175 (2010)
14. Ceragioli, F., De Persis, C., Frasca, P.: Quantized average consensus: discontinuities and hysteresis. *Automatica* (to appear)
15. Dimarogonas, D.V., Johansson, K.H.: Stability analysis for multi-agent systems using the incidence matrix: quantized communication and formation control 46, 695–700 (2010)
16. Li, T., Fu, M., Xie, L., Zhang, J.F.: Distributed consensus with limited communication data rate. *IEEE Trans. Automat. Contr.* 56(2), 279–292 (2011)

# Research on Automotive Parts Abradability on Driving Behavior Analysis<sup>\*</sup>

Hu Wang and Xuan Zhang

School of Management  
Wuhan University of Technology  
Wuhan, P.R. China

wanghu61@126.com, zxxg1111@163.com

**Abstract.** This paper analyzed the impact of driver's behaviors on automotive parts performance, and found the appropriate indicator for measurement of wear degree of automotive parts. What's more, the quantitative calculation model of wear degree of automotive parts on driving behavior analysis was presented in this article. The model is useful to study for offering the different initiative service to customers.

**Keywords:** driver's behavior, analysis different initiative service, automotive parts wear, wear numerical calculation.

## 1 Introduction

Seeing from the current conditions of automotive after-sales service, most customers repair their cars when there is something wrong with their automotive parts. Usually, customers only have their cars serviced regularly following the normal maintenance. So, automotive after-sales service still stay in passive service. Against the weakness in automotive after-sales service [1], I proposed a model which can improve the efficiency and the quality of automotive after-sales service. The core idea of the model is to predict the car's troubles and the time when the troubles will occur through analyzing the impact of driver's behavior on automotive parts performance.

According to the survey, there are many factors which have different effects on automotive parts performance, such as road conditions route, traffic, maintenance period etc, and the influence of the factors can be found by analyzing driver's behavior. So calculate the wear degree of auto parts which is caused by driver's behavior firstly. Then predict the car's troubles and the time when the troubles will occur by contrasting the useful life of the auto parts [2].

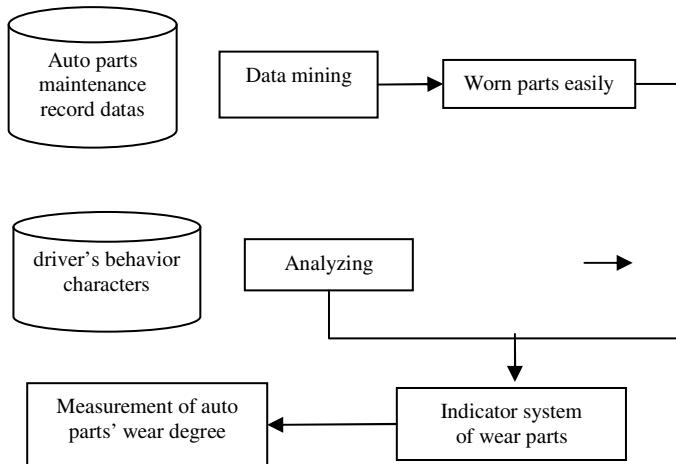
## 2 Quantitative Calculation Model

The system of a car consists of many sub-systems which contain assembly parts, and the bases of them are parts. So cars' failures are usually described by automotive parts'

\* 2010, Natural Sciences Foundation of China, project number: 71071122. Hubei Science and Technology Agency Research, project number 20102s0014.

failures. During the car is working, automotive parts wear is affected by external environment driving conditions. Finally, the parts lost their function. The index system of parts wear could be built by analyzing driver's behavior and driving conditions, and then calculates the wear degree of auto parts through the index system.

According to above-mentioned, the indicators of parts wear are selected by analyzing driver's behaviors and driving conditions, and then calculate the wear degree of auto parts through the index system, showing in Fig. 1.



**Fig. 1.** Model of automotive parts wear degree

### 3 Auto Parts Abradability Algorithm

#### 3.1 Analysis the Easy Worn Parts

According to analyzing the structure of the car, the whole system can be divided into following parts: engine, chassis, automotive body and electric apparatus. The parts contain 17 assemblies such as car body, oil supply system, clutch, display instrument etc. So analyze the interaction between all parts firstly. According to the maintenance records, questionnaires and expert knowledge, and then get the easy wear parts in the car working. The number of each part is  $P_j$  ( $j=1,2 \cdots n$ ,  $n$  is the parts count) , showing in Table 1.

**Table 1.** Easy-wear parts

$P_1$	$P_2$	...	$P_j$	...
Engine	Tire	...	Clutch	...

### 3.2 Indicator System of Parts Wear

The main factors for effects on wear parts are the external environment conditions of driving and driver's distinctive behaviors characters. The objective conditions of driving include: road surface, traffic, climatic etc.

The behavior characters contain: speed of driving, skill, mileage, driving behavior, maintenance period etc. According to the questionnaires and expert knowledge, I select the impact-index from the factors [4]. The impact-index of each part wear is  $F_{ij}$  (the  $i$ th impact-index of  $j$ th part wear). The weight value of  $F_{ij}$  is a  $a_{ij}$  ( $j=1,2 \cdots n$ ;  $i=1,2 \cdots m$  suppose the indexes number is  $m$ ) and  $\sum_{i=1}^m a_{ij} = 1$ , showing in table 2.

**Table 2.** Symbol of indicator and weigh

Indicator	Weight
$F_{1j}$	$a_{1j}$
$F_{2j}$	$a_{2j}$
$F_{3j}$	$a_{3j}$
...	...
$F_{mj}$	$a_{mj}$

By analyzing driver's behaviors preferences, select the common impact-factors (CF) of driver's behaviors on wear parts and the personal factors(PF). CF is the factors which driver's distinctive behaviors have little influence on wear parts. By contrast, PF have significant effect on wear parts.

### 3.3 Wear Numerical Calculation

Based on customer classification, the impact indicators are divided into common indicators and personal indicators according to customers generality characteristic and features. Study on the law and way that each indicator affects on wear parts firstly, such as the wear degree of tire is linear to the route of driving. Then build the mathematical model, and calculate the wear degree of each indicator on parts wear  $w_{ij}$  (the wear degree of indicator  $I$  on part  $j$ ;  $i=1,2 \cdots n$ ;  $j=1,2 \cdots m$ ).

Personal impact-indicators describe the driver's individual characters. In order to show the differences between individual behavior and standard behavior, the influence coefficient  $\beta_{ij}$  is given by experts according to driver's behavioral

differences. So the wear degree of influence indicator on driver's individual behavior  $w_{ij} = w_{ij} \cdot \beta_{ij}$  (suppose the count CF is h; PF is n-h), showing in table 3.

**Table 3.** Parts' wear degree

<i>Indicator</i>	<i>Part</i> <i>Indicator</i>	P <sub>1</sub>	P <sub>2</sub>	...	P <sub>j</sub>	...
CF <sub>1</sub>		W <sub>11</sub>	W <sub>12</sub>	...	W <sub>1j</sub>	...
CF <sub>2</sub>		W <sub>21</sub>	W <sub>22</sub>	...	W <sub>2j</sub>	...
...		...	...	...	...	...
CF <sub>h</sub>		W <sub>h1</sub>	W <sub>h2</sub>	...	W <sub>hj</sub>	...
PF <sub>h+1</sub>		P <sub>j</sub>	W <sub>h+12</sub>	...	W <sub>h+1j</sub>	...
...		...	...	...	...	...

## 4 Example of Tire Abradability

According to the algorithm of automotive parts wear degree, take the tire for example; evaluate the abrasion value W<sub>21</sub> of different road surface caused by driver's behavior. Tire wear is a system and complex dynamic behavior, so it shows academic and practical importance for suspension research and tire structure improvement.

### 4.1 Indicator System of Tire Wear

The process of abrasion on tire is very complicated, and the wear is many mechanism result of joint action. The main factors are the quality of tire, matching of complete vehicle, and external environment, such as the structure of tire, abrasion resistance of tire tread, atmospheric pressure; load, speed, traffic effort, suspension. What's more, the abrasion has directed relationship with road surface, temperature and operative operation.

This paper attention is concentrated on the tire wear caused by external environment. According to questionnaire and expert knowledge, I select traffic, road surface, route, driving skill, driver's behavior and maintenance period as the measurement index of tire abrasion. Suppose the drivers are in the same city, using the same cars, then the common factors and personal factors which are caused to tire wear are divided following: CF are traffic, road surface; PF include driving skill, route, driving behavior and maintenance period, showing in table 4.

**Table 4.** Measurement index of tire wear

<i>Index</i>	<i>Weight</i>
CF <sub>21</sub> traffic	a <sub>21</sub>
CF <sub>22</sub> road surface	a <sub>22</sub>
PF <sub>23</sub> driving skill	a <sub>23</sub>
PF <sub>24</sub> driving behavior	a <sub>24</sub>
PF <sub>25</sub> maintenance period	a <sub>25</sub>

#### 4.2 Analysis Tire Tribology Theory

With stability theory of dynamic system, integrated by randomicity of tire tread vibration and location on circle direction, the quantitative analysis of tire wear is studied to tread unstable vibration.

- 1) Homogeneous Wear: When car is working without disturbance, the tire is worn evenly by road. Under this circumstance, the car system will reach relatively steady state. So suppose the tire wear caused by the system quality when tire has undisturbed.
- 2) Disturbed Wear: The tire circumference exists the heavy wear area because the tire is disturbed. So tire wear is calculated by the superposition of homogeneous and disturbed wear. From the abrasion's peak value, the wear is recurring and regularly. The state is polygon wear. The sides and speed are related to frequency of tire disturbed. The abrasion is related to wear speed and coefficient between tires.

#### 4.3 Predication of Wear Abrasion

During the course of driving, the tire will subject to the real-time affect from the environment, vehicle self, tire and driver's behavior which lead to forming different texture in the tread.

Through defining a based road sample and the proportion of composed sub-roads in the base sample, tire life is forecasted by using tire wear linearly additive method. Then the method on forecasting tire wear is built compressively and detailedly.  
*1)Road Surface:* According to the research, the data of road proportion and driving speed in tire life cycle are as shown in table 5:

**Table 5.** Road proportion and driving speed

<i>Typical Road</i>	<i>Highway</i>	<i>City road</i>	<i>Country Road</i>	<i>Bad Road</i>
Speed	100-29 (km/h)	50-60 (km/h)	40-60 (km/h)	30-40 (km/h)
Proportion	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>

2) Road Surface Wear Coefficient: During the car is driving, the friction between road surface and tire is the main reason of tire wear. The friction coefficients are different in different road which are related to tire material and roughness of road. The friction coefficient grows small, and tire is easy to relate slid on the road. The wear coefficient is bigger. So suppose the friction coefficients between road surface and tire inversely proportional to wear coefficient, then we can get the Homogeneous wear coefficient and Disturbed wear coefficient from friction peak value, showing in table 6.

**Table 6.** The wear coefficient between road and tire coefficient for the typical road attached by tires[5]

<i>Typical Road</i>	<i>Highway</i>	<i>City rod</i>	<i>County Road</i>	<i>Bad Road</i>
Friction Peak Value	0.9-0.8	0.7-0.8	0.5-0.6	0.68
Averagely wear coefficient	1e-13	1.08e-13	1.68e-13	1.94e-13
Disturbed wear coefficient	1.2e-12	1.3e-12	1.7e-12	2.1e-12

#### 4.4 Wear Numerical Calculation

Tire wear is calculated by the route in different road surface. So the road surface and route indicators can be merged to calculate the tire abrasion. Calculate the tire polygon wear on different road surface firstly. Then superimpose the tread circumference. Finally get the abrasion caused by route on different road surface.

According to the periodically of tire rolling, add the height of tire wear to tread circumferential position, then the abrasion of tire on circumferential position is calculated. Suppose the step in iteration is  $\Delta t$ , the cycle of rolling is  $T$ , then the number of wear point in the rolling cycle is:

$$D = \frac{T}{\Delta t} \quad (1)$$

The total wear cased by road is:

$$W_{road} = W_{averagely} + W_{disturbed} \quad (2)$$

The formula of tire Wear numerical calculation caused by road condition is as following:

$$W_{21} = P_1 W_{highway} + P_2 W_{cityroad} + P_3 W_{countryroad} + P_4 W_{badroad} \quad (3)$$

According to the method, we can calculate the tire abrasion, and predict the car's troubles and the time by contrasting the useful life of the auto parts. Based on that, we could offer the different initiative service to customers.

## 5 Conclusion

In auto after-sales service, the service model which is offering the different initiative service to customers is in the beginning stage. According to analyzing the car wear caused by driver's behavior, this paper predicts the car's troubles and the time when the troubles will occur by contrasting the useful of the auto parts. What's more, this article builds the index system, and presents the algorithm of auto wear numerical calculation. However, this article is only useful to study for offering the different initiative service to customers. So it needs to be further research and improvement.

## References

1. Wang, H., Yuan, H.: Service-mining Based on Customer Value Analysis. In: 2007 International Conference on Management Science & Engineering (14th), August 20-22, pp. 109–114 (2007)
2. Wang, H., Yu, L.: Research on the model of service mining under active service-oriented. Wuhan University of Technology (Information Management and Engineering) 32(2) (2010)
3. Pohl, A., Steindl, R., Reindl, L.: The 'Intelligent Tire' Utilizing Passive SAW Sensors-Measurement of Tire Friction. IEEE Transaction on Instrument and Measurement 6(48), 1041–1046 (1999)
4. Yilmaz, V., Celik, H.: A model for risky driving attitudes in Turkey. Social Behavior and Personality 32(8), 791–796 (2004)
5. Yu, S.Z.: Auto-Theories (the third editor), pp. 74–75. Hong Xie Tech Publication, Beijing (2000)
6. Tian, L., Zhi, F.H., Chen, D.Z.: Customers Based on Classification Method of Fuzzy Neural Network. Productivity 12, 120–121 (2007)
7. Li, Y.Y., Yun, J.: Multiple Attribute Summary of Comprehensive Evaluation Index System Theory. Wuhan University of Technology Information and Management Engineering 2(31), 305–309 (2009)
8. Lachmbe, J.: Tire model for simulation of vehicle motion on high and low friction road surface. In: Proceeding of 2000 Winter Simulation Conference, pp. 1025–1034 (2000)
9. Wahab, T.G., Wen, K.N.: Technol. Understanding driver behavior using multi-dimensional CM CMAC. In: 2007 International Conference on Information, Communications & Signal Processing (6th), December 10-13, pp. 1–5 (2007)
10. Sevrndenius, J.: Wheel, model review and friction estimation. Technical Report ISRN Association, Cambridge, UK (2003)

# Existence and Stability of Equilibrium of Discrete-Time Neural Networks with Distributed Delays

Xuyang Lou<sup>\*</sup>, Baotong Cui, and Qian Ye

Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education),

Jiangnan University, Wuxi 214122, China

Xuyang.Lou@gmail.com

**Abstract.** We study the global stability and existence of discrete-time neural networks with distributed delays by using semi-discretization technique and a Lyapunov functional. A sufficient condition is presented in this paper for the existence and global exponential stability of the equilibrium point of discrete-time neural networks with distributed delays. It is shown that the convergence dynamics of these kind of continuous-time networks reported so far in the literature are preserved by the discrete-time analogues without any restriction on the discretization step-size.

**Keywords:** Neural networks, discrete-time, distributed delays.

## 1 Introduction

Neural networks with time-delay are often used to describe dynamic systems due to its practical importance and wide applications in many areas such as industry, biology, economics and so on. In such applications, it is of prime importance to ensure that the equilibrium points of the designed network is stable. So, it is important to study the stability of neural networks. Recently, many authors have studied the existence and stability of the following delayed neural networks (DNNs) [1]-[3]:

$$\frac{dx_i(t)}{dt} = -a_i x_i(t) + \sum_{j=1}^m b_{ij} f_j(x_j(t)) + \sum_{j=1}^m c_{ij} f_j(x_j(t - \tau_{ij})) + u_i, \quad (1)$$

where  $x_i(t)$  ( $i = 1, 2, \dots, m$ ) are the activations of the  $i$  th neurons;  $a_i > 0$  ( $i = 1, 2, \dots, m$ ) are positive constants, they denote the rate with which the cell  $i$  reset their potential to the resting state when isolated from the other cells and inputs;  $b_{ij}, c_{ij}$  ( $i, j = 1, 2, \dots, m$ ) are the connection weights, they denote the strengths of connectivity between the cells  $i$  at time  $t$  and  $t - \tau_{ij}$  ( $i, j = 1, 2, \dots, m$ ), respectively; time delays  $\tau_{ij}$  ( $i, j = 1, 2, \dots, m$ ) are nonnegative constants, which

---

\* This work is partially supported by National Natural Science Foundation of China (No.61174021, No.61104155), the 111 Project (B12018), and the Jiangsu Provincial Program for Postgraduate Scientific Innovative Research of Jiangnan University (No. CXZZ11\_0463).

correspond to the finite speed of the axonal signal transmission;  $u_i$  ( $i = 1, 2, \dots, m$ ) denote the  $i$  th component of an external input source introduced from outside the network to the cell  $i$ . For system (1), we assume that

$$(H_1) |f_i(\cdot)| \leq M_i, i = 1, 2, \dots, m, x \in \mathbb{R} \text{ for some constant } M_i > 0.$$

( $H_2$ ) There exists a positive number  $L_i$  such that

$$|f_i(x) - f_i(y)| \leq L_i |x - y|,$$

for all  $x, y \in \mathbb{R}, i = 1, 2, \dots, m$ .

In recent years, some authors [4]-[6] pay attention to neural networks with distributed delays. In numerical simulations and practical implementation of a continuous-time neural network, discretization is needed. Therefore, it is of both theoretical and practical importance to study the dynamics of discrete-time neural networks. When a continuous-time dynamical system is discretized for computer simulations and computational purposes, it is usually expected that the dynamical characteristics of the continuous-time system pass to its discrete-time analogue. While there are numerous ways of obtaining discrete-time analogues from their continuous-time dynamical systems, most of the discrete-time analogues do not faithfully preserve the dynamics of their continuous-time versions (see [7], for instance). We refer to [8] for an emphasis on the need for discrete-time analogues to reflect the dynamics of their continuous-time counterparts. In [9], Li studies the existence and stability of periodic solutions for the discrete neural networks with delays. Based on the linear matrix inequality (LMI), Liang et al [10] present some sufficient conditions for the existence, uniqueness and global exponential stability of the equilibrium point of discrete-time bidirectional associative memory (BAM) neural networks with variable delays. Convergence dynamics of continuous-time and discrete-time bidirectional neural networks with constant transmission delays are studied in [11]. However, up to now, few authors have studied the dynamical behaviors of the discrete-time analogues of neural networks with distributed delays [12].

Motivated by the above discussion, we focus on the existence and global exponential stability of the equilibrium point of discrete-time neural networks with distributed delays and derive sufficient conditions. The convergence dynamics of the class of continuous-time counterparts are preserved by the discrete-time analogues without any restriction on the discretization step-size.

## 2 Discrete-Time Analogues

The neural networks with distributed delays in continuous-time are described as follows

$$\begin{aligned} \frac{dx_i(t)}{dt} = & -a_i x_i(t) + \sum_{j=1}^m b_{ij} f_j(x_j(t)) + \sum_{j=1}^m c_{ij} f_j(x_j(t - \tau_{ij})) + u_i \\ & + \sum_{j=1}^m d_{ij} f_j \left( \int_0^\infty K_{ij}(s) x_j(t-s) ds \right), \end{aligned} \quad (2)$$

for  $i = 1, 2, \dots, m$ , where  $f_i(\cdot)$  ( $i = 1, 2, \dots, m$ ) also satisfy  $(H_1)$  and  $(H_2)$ , the delay kernels  $K_{ij}(\cdot)$ ,  $i, j = 1, 2, \dots, m$  are assumed to satisfy the following conditions:

- (i)  $K_{ij} : [0, \infty) \rightarrow [0, \infty)$ ;
- (ii)  $K_{ij}$  is bounded and continuous on  $[0, \infty)$ ;
- (iii)  $\int_0^\infty K_{ij}(s)ds = 1$ ;
- (iv) there exists a positive number  $\mu$  such that

$$\int_0^\infty K_{ij}(s)e^{\mu s}ds < \infty.$$

First, we replace the integral term

$$\int_0^\infty K_{ij}(s)x_j(t-s)ds, i, j = 1, 2, \dots, m, \quad (3)$$

by an approximate discrete sum. Let  $h > 0$  be a fixed real number. Let  $[t/h]$  and  $[s/h]$  denote the greatest integers contained in  $t/h$  and  $s/h$ , where  $t$  and  $s$  denote arbitrary real nonnegative numbers. The above integral is approximated by a sum of the form

$$\begin{aligned} & \sum_{[s/h]=1}^{\infty} w_{ij}(h)K_{ij}\left(\left[\frac{s}{h}\right]h\right)x_j\left(\left[\frac{t}{h}\right]h - \left[\frac{s}{h}\right]h\right) \\ &= \sum_{[s/h]=1}^{\infty} k_{ij}\left(\left[\frac{s}{h}\right]h\right)x_j\left(\left[\frac{t}{h}\right]h - \left[\frac{s}{h}\right]h\right), \end{aligned}$$

where  $w_{ij}(h)$  denote positive weight functions which can be chosen appropriately so that the resulting discrete kernels  $k_{ij}(\cdot)$  will satisfy some properties given in (12) below. For convenience, we let  $[t/h] = n$  and  $[s/h] = p$  where  $p$  and  $n$  denote integers. We also use the notation,  $k_{ij}(ph) = k_{ij}(p)$  and  $x_j(nh - ph) = x_j(n - p)$  for  $i, j = 1, 2, \dots, m$ . With these preparations, we reformulate (2) by equations with piecewise constant arguments of the form

$$\begin{aligned} \frac{dx_i(t)}{dt} &= -a_i x_i(t) + \sum_{j=1}^m b_{ij} f_j\left[x_j\left(\left[\frac{t}{h}\right]h\right)\right] + \sum_{j=1}^m c_{ij} f_j\left[x_j\left(\left[\frac{t}{h}\right]h - \left[\frac{\tau_{ij}}{h}\right]h\right)\right] \\ &+ \sum_{j=1}^m d_{ij} f_j \left[ \sum_{[s/h]=1}^{\infty} K_{ij}\left(\left[\frac{s}{h}\right]h\right)x_j\left(\left[\frac{t}{h}\right]h - \left[\frac{s}{h}\right]h\right) \right] + u_i, \end{aligned} \quad (4)$$

for  $i = 1, 2, \dots, m, t \in \left(\left[\frac{t}{h}\right]h, \left[\frac{t}{h}\right]h + h\right)$ , where  $h$  is a positive number denoting a uniform discretization step size and  $[t/h]$  denotes the integer part of  $t/h$ . For convenience, in the following, we let  $x_i(nh) = x_i(n)$ ,  $[t/h] = n$  and  $[\tau_{ij}/h] = r_{ij}$ , where  $r_{ij}$  are nonnegative integers. Then, the system (2) can be put in the form

$$\begin{aligned} \frac{dx_i(t)}{dt} = & -a_i x_i(t) + \sum_{j=1}^m b_{ij} f_j(x_j(n)) + \sum_{j=1}^m c_{ij} f_j(x_j(n-r_{ij})) \\ & + \sum_{j=1}^m d_{ij} f_j\left(\sum_{p=1}^{\infty} k_{ij}(p) x_j(n-p)\right) + u_i, \end{aligned} \quad (5)$$

where  $i = 1, 2, \dots, m, t \in [nh, (n+1)h], n \in Z^+, Z^+ = \{1, 2, \dots\}$ . Let

$$\Omega = \sum_{j=1}^m b_{ij} f_j(x_j(n)) + \sum_{j=1}^m c_{ij} f_j(x_j(n-r_{ij})) + \sum_{j=1}^m d_{ij} f_j\left(\sum_{p=1}^{\infty} k_{ij}(p) x_j(n-p)\right) + u_i.$$

Since

$$\frac{d}{dt}(x_i(t)e^{a_i t}) = -e^{a_i t} \Omega. \quad (6)$$

We integrate (6) over the interval  $[nh, t]$ , where  $t < (n+1)h$ , and obtain

$$x_i(t)e^{a_i t} - x_i(n)e^{a_i nh} = \frac{e^{a_i t} - e^{a_i nh}}{a_i} \Omega. \quad (7)$$

Allowing  $t \rightarrow (n+1)h$  and simplifying, we obtain a discrete-time analogue of (2) written as

$$x_i(n+1) = e^{-a_i h} x_i(n) + \frac{1 - e^{-a_i h}}{a_i} \Omega, \quad (8)$$

for  $i = 1, 2, \dots, m, t \in [nh, (n+1)h], n \in Z_0, Z_0 = \{0, 1, 2, \dots\}$ .

An equilibrium of (8) is given by  $x^* = (x_1^*, x_2^*, \dots, x_m^*)^T$ , where

$$\frac{1 - e^{-a_i h}}{a_i} \left\{ a_i x_i^* - \left( \sum_{j=1}^m b_{ij} f_j(x_j^*) + \sum_{j=1}^m c_{ij} f_j(x_j^*) + \sum_{j=1}^m d_{ij} f_j\left(\sum_{p=1}^{\infty} k_{ij}(p) x_j^*\right) + u_i \right) \right\} = 0. \quad (9)$$

Since  $a_i > 0$  and  $h > 0$  imply  $\frac{1 - e^{-a_i h}}{a_i} > 0$ , it follows that the equilibria of (2) and (8) coincide.

For convenience, we let

$$\phi_i(h) = \frac{1 - e^{-a_i h}}{a_i}, i = 1, 2, \dots, m,$$

and note that  $\phi_i(h) > 0$  when  $a_i > 0$  and  $h > 0$ . So (8) becomes

$$\begin{aligned} x_i(n+1) = & e^{-a_i h} x_i(n) + \phi_i(h) \sum_{j=1}^m b_{ij} f_j(x_j(n)) + \phi_i(h) \sum_{j=1}^m c_{ij} f_j(x_j(n-r_{ij})) \\ & + \phi_i(h) \sum_{j=1}^m d_{ij} f_j\left(\sum_{p=1}^{\infty} k_{ij}(p) x_j(n-p)\right) + \phi_i(h) u_i, i = 1, 2, \dots, m, n \in Z_0. \end{aligned} \quad (10)$$

We suppose that (8) is supplemented with initial values of the form

$$x_i(l) = \varphi_i(l), l \in (-\infty, 0], i = 1, 2, \dots, m, \quad (11)$$

where  $l \in (-\infty, 0]$  denotes  $l \in \{\dots, -2, -1, 0\}$  and the sequence  $\{\varphi_i(l)\}_{l=-\infty}^{l=0}$  is bounded for all  $i = 1, 2, \dots, m$ .

We assume that the discrete kernels  $k_{ij}(\bullet)$  for  $i, j = 1, 2, \dots, m$  satisfy the following conditions:

(i')  $k_{ij}(p) \in [0, \infty)$  ( $p = 1, 2, 3, \dots$ ) and  $k_{ij}(p)$  is bounded;

(ii')  $\sum_{p=1}^{\infty} k_{ij}(p) = 1$ ;

(iii') there exists a number  $v > 1$  such that

$$\sum_{p=1}^{\infty} k_{ij}(p)v^p < \infty. \quad (12)$$

With the above properties, we observe the following: let  $v = e^{\mu}$ , where  $\mu > 0$ , and let  $p \in \{1, 2, 3, \dots\}$ . For any number  $1 < \lambda < v$  and by letting  $\lambda = e^{\delta}$ , where obviously  $0 < \delta < \mu$ , we have  $p\lambda^p = pe^{\delta p} < \exp\left(\frac{\mu p}{2}\right)\exp\left(\frac{\mu p}{2}\right)$  for large  $p$ .

As a result, we get

$$\sum_{p=1}^{\infty} pk_{ij}(p)\lambda^p = \sum_{p=1}^{\infty} pk_{ij}(p)e^{\delta p} < \sum_{p=1}^{\infty} k_{ij}(p)e^{\mu p} < \infty. \quad (13)$$

### 3 Main Results

In the following result, we establish an easily verifiable set of sufficient conditions for the existence of a equilibrium for system (10) and the global exponential stability of the equilibrium  $x^*$ .

**Lemma 1:** Suppose  $(H_1)$  holds. Then, the equilibrium  $x^*$  of (4) or (10) is existent.

**Theorem 1:** Let  $h > 0$  and let the discrete kernels  $k_{ij}(\bullet)$ ,  $i, j = 1, 2, \dots, m$  satisfy (12). Suppose the conditions  $(H_1)$  and  $(H_2)$  hold. Then, if the system parameters satisfy  $a_i > L_i \sum_{j=1}^m (|b_{ji}| + |c_{ji}| + |d_{ji}|)$ ,  $i = 1, 2, \dots, m$ , there exist constants  $\alpha \geq 1$  and  $\lambda > 1$  such that all solutions of (10) satisfy

$$\sum_{i=1}^m \frac{|x_i(n) - x_i^*|}{\phi_i(h)} \leq \alpha \left( \frac{1}{\lambda} \right)^n \sum_{j=1}^m \left( \sup_{l \in (-\infty, 0]} \frac{|\varphi_j(l) - x_j^*|}{\phi_j(h)} \right)$$

where  $n \in Z^+$  and  $l \in (-\infty, 0]$  denotes  $l \in \{\dots, -2, -1, 0\}$ .

## 4 Conclusion

In this paper, we obtain a discrete-time analogue of system (2) by means of semi-discretization technique and establish sufficient conditions for the existence and the global exponential stability of the equilibrium point of discrete-time neural networks with distributed delays. One numerical example is given to illustrate the effectiveness of the obtained results.

## References

1. Joy, M.: Results concerning the absolute stability of delayed neural networks. *Neural Networks* 13, 613–616 (2000)
2. Liao, X.F., Chen, G., Sanchez, E.N.: LMI-based approach for asymptotically stability analysis of delayed neural networks. *IEEE Trans. Circuits Syst. I* 49(7), 1033–1039 (2002)
3. Singh, V.: A generalized LMI-based approach to the global asymptotic stability of delayed cellular neural networks. *IEEE Trans. Neural Networks* 15, 223–225 (2004)
4. Zhou, J., Li, S.Y., Yang, Z.G.: Global exponential stability of Hopfield neural networks with distributed delays. *Appl. Math. Model.* 33(3), 1513–1520 (2009)
5. Lu, J.G.: Robust global exponential stability for interval reaction-diffusion Hopfield neural networks with distributed delays. *IEEE Trans. Trans. Circuits Syst. II-Express Briefs* 54(12), 1115–1119 (2007)
6. Oliveira, J.J.: Global asymptotic stability for neural network models with distributed delays. *Math. Comput. Model.* 50(1-2), 81–91 (2009)
7. Blum, E.K., Wang, X.: Stability of Fixed-Points and Periodic-Orbits and Bifurcations in Analog Neural Networks. *Neural Networks* 5(4), 577–587 (1992)
8. Stuart, A.M., Humphries, A.R.: *Dynamical Systems and Numerical Analysis*. Cambridge University Press, Cambridge (1996)
9. Li, Y.K.: Global stability and existence of periodic solutions of discrete delayed cellular neural networks. *Phys. Lett. A* 333(1-2), 51–61 (2004)
10. Liang, J.L., Cao, J.D., Ho, D.W.C.: Discrete-time bidirectional associative memory neural networks with variable delays. *Phys. Lett. A* 335(2-3), 226–234 (2005)
11. Mohamad, S.: Global exponential stability in continuous-time and discrete-time delayed bidirectional neural networks. *Physica D* 159(3-4), 233–251 (2001)
12. Mohamad, S.: Exponential stability preservation in discrete-time analogues of artificial neural networks with distributed delays. *J. Compu. Appl. Math.* 215(1), 270–287 (2008)

# Construction Scheme and Key Technologies of Electric Energy Information Acquisition System

Enguo Zhu and Xuan Liu

China Electric Power Research Institute  
100192 Beijing, China  
enguozhu@163.com

**Abstract.** Electric energy information acquisition system is the physical basis of China strong and smart grid in power consumption field. It analyzes the needs of construing smart grid for electric energy information acquisition system and concludes its present construction status. The construction mode and technology scheme of acquisition system is presented. Different electric energy information acquisition modes are proposed for different power users. It analyzes power line carrier communication standardization, security protection of acquisition system, pre-paid control and other key technologies. The system can greatly speed up standardization construction of marketing measurement, meter reading, charging and information system of electric power company. It can further enhance marketing business management level and provide important technology guarantee for achieving step tariff and speeding up construction of strong and smart grid.

**Keywords:** electric energy information acquisition system, construction scheme, smart grid, security protection, communication standardization, prepay fee control.

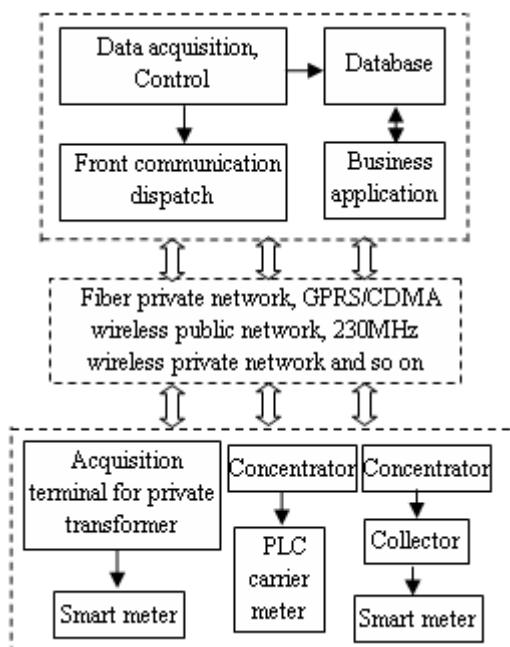
## 1 Introduction

Smart grid uses communication, advanced sensor and distributed computing technology to operate and control power system. It cannot only improve the energy structure and utilization efficiency but also improve the economy, security and reliability of power transmission system [1-4]. With the rapid development of science and technology and the increment of information management level, electric power companies have built many load management systems and concentrated meter reading systems, which play an important role in electric power security production. But these acquisition systems are small, scattered, isolated and the overall collection coverage ratio is very low. They cannot come into being an integrated acquisition platform covering all the power users. They have many types of equipments and communication methods but there is no uniform system construction standard. The comprehensive data utilization level of these acquisition systems is very low and cannot meet the needs of all the levels and professions.

How to improve the automation acquisition level of electric energy information and the accuracy and timeliness of the line loss analysis is of great importance to establish power consumption scheme, ensure the safety and stability of power grid, improve service capability and meet the diverse needs of customers[5]-[6]. Electric energy information acquisition system is the physical basis of constructing smart grid. It uses advanced sensor, communication, automatic control and other technologies to acquire and manage data, analyze power quality and line loss statistical data. It can timely acquire electric energy information, find abnormal power consumption, monitor and control electricity load. It can provide physical basis for achieving step tariff, pre-paid fees and other marketing service strategies.

## 2 The Overall Structure of Electric Energy Information Acquisition System

As shown in Fig.1, the electric energy information acquisition system is composed by main station, communication channel and acquisition devices.



**Fig. 1.** Composition schematic diagram of electric energy information acquisition system

## 2.1 Main Station

Main station is composed by database server, disk array, application server, front-end server, interface server, workstation, GPS clock, firewall and other related network equipments. It is used to complete business application, data acquisition, control, front-end communication dispatch, database management and other functions.

## 2.2 Communication Channel

Communication channel includes fiber private network, GPRS/CDMA wireless public network, 230MHz wireless private network and so on. It is used to transfer data from acquisition devices to main station.

## 2.3 Acquisition Devices

Acquisition devices include acquisition terminal for special transformer, concentrator, collector, smart meter and so on. They are installed to collect and provide initial power consumption information of the whole system.

# 3 Acquisition Mode of Electric Energy Information Acquisition System

The electric energy information acquisition system can acquire all the power users' electric energy information. There are many power users in China and their electrical environments are different. So they use different types of remote communication channels and acquisition terminals. According to the different characteristics of power users, the electric energy information acquisition system uses different technology scheme to collect and monitor power consumption information of all the power users.

## 3.1 Large-Scale Private Transformer User

The large-scale private transformer user installs acquisition terminal with load management function, which can timely acquire the output pulse by 485 bus, obtain and store measurement data and information of smart meter. It can complete load control function by monitoring the switch of power user, transfer power consumption information to main station by remote communication channel and provide local information service for power user.

## 3.2 Small and Medium Scale Private Transformer User

The small and medium scale user installs acquisition terminal for private transformer, which can communicate with smart meter by 485 bus, obtain and store measurement data and information of smart meter. It can complete pre-paid control and management by monitoring the switch of power user, transfer power consumption information to main station by remote communication channel, receive and perform the management of main station.

### 3.3 Low-Voltage Single or Three Phase Industrial or Commercial User

The technology scheme can complete remote meter reading, timely acquire power consumption information and analyze abnormal data. Under the supervision of main station, it can achieve prepaid management function by directly controlling smart meter.

### 3.4 Residential User and Public Power Distribution Transformer

There are a huge number of urban and rural users in China and the electric energy information of a single user is little. The data acquisition network is composed by concentrator, collector, smart meter and data transmission channel. It is used to complete remote meter reading and power consumption supervision. The concentrator transmits electric energy information to main station by fiber private network or GPRS/CDMA wireless public network and complete prepaid management.

Each public power distribution transformer is an acquisition unit for achieving meter reading of residential customers. The concentrator acquires all the data of smart meter by local communication channel.

## 4 The Key Technology of Electric Energy Information Acquisition System

### 4.1 Standardization of Power Line Communication

Power line communication can extend communication network to low-voltage user side instead of laying special communication line. It is very suitable for the area that has scattered smart meter, difficult to lay wire and little change in load characteristics. But the communication method has shortcomings such as signal attenuation, many noise sources, strong interference, easy to be affected by load characteristics and low communication reliability.

Currently, the low-voltage power line communication technology applications have different technical characteristics. They are not benefit for resident meter reading in the construction of electric energy information acquisition system and restrict the development of the carrier wave communication technology. Therefore, it is very necessary to research the feasibility of communication standardization technology of acquisition devices. According to the main technical characteristics and parameters such as center frequency, modulation method, spread frequency code length, occupied bandwidth, the maximum transmission level, receiver sensitivity, variable load impedance, transmission protocol, router algorithm, related technical test and parameter measurement method, the technical standard should be presented to make acquisition devices join, connect and exchange each other. So it can regulates device manufacture, engineering implement and ensure the construction of electric energy information acquisition system and strong smart grid.

## 4.2 Security Protection Technology of Acquisition System

With the rapid development of information technology in power industry, electric energy information acquisition system uses a lot of communication, network and other new technology. Though the information automation level is increased, there are more and more security risks in acquisition system. The information security problem is very important because it will threaten security, stability and economy operation of power system. The electric energy information acquisition system is one of the most core business application systems. The information of customer and electric energy is the core data resource of electric power company and involves all aspects of social life. Illegal using or leaking information will bring irreparable damage.

Now it is very necessary to research the security protection status of electric energy information and management system, analyze and assess the potential security problems in all the aspects of acquisition system, propose construction program and construct an effective security protection system. According to the overall requirements of constructing security protection system, it should take measures from main station, acquisition devices and communication channel to ensure the security of acquisition system and improve its total security protection capability.

## 4.3 Pre-paid Control Technology of Acquisition System

With the improvement of power information level and development of marketing business, electric energy information acquisition system uses comprehensive pre-paid power consumption management mode. Customers firstly pay electricity bill to ensure their power consumption. The acquisition system acquires electric energy information, computes remain fee and displays it to customers. If there is no much fee, it will remind customers to pay electricity bill and perform break off power consumption when there is no fee. Pre-paid management can be performed by main station, acquisition terminal and smart meter. The according control method includes main station, acquisition terminal and smart meter pre-paid. Their application areas are shown as Table 1.

**Table 1.** Comparison of three pre-paid control modes

Pre-paid method	Main station pre-paid	Acquisition terminal pre-paid	Smart meter pre-paid
Logic perform place	Main station perform pre-paid control logic	Acquisition terminal perform pre-paid control logic	Smart meter perform pre-paid control logic
Application area	Private transformer user, the single or three phase industrial or commercial user, residential customer	Private transformer user, but has large error	Residential customer, reduce pressure of main station with huge number of users

## 5 Conclusions

Electric energy information acquisition system can provide complete and accurate data support for SG186 business and improve automation level of energy measurement, automatic meter reading, prepayment fees and other marketing business. Currently, according to the uniform technical standard and scheme, SGCC is constructing the acquisition system. Up to now, more than 50 million user's power consumption information can be automatically acquired and the system is in good operation. Its application can meet the urgent requirements of all levels and specialties for power consumption information and has good economic and social benefits.

## References

1. Wang, C., Li, P.: Development and Challenges of Distributed Generation, the Micro-grid and Smart Distribution System. *Automation of Electric Power Systems* 34(2), 10–14 (2010)
2. Zhang, W., Liu, Z., Wang, M.: Research Status and Development Trend of Smart Grid. *Power System Technology* 33(13), 1–11 (2009)
3. Chang, K., Xue, F., Yang, W.: Review on the Basic Characteristics and its Technical Progress of Smart Grid in China. *Automation of Electric Power Systems* 33(17), 10–15 (2009)
4. Xiao, S.: Consideration of Technology for Constructing Chinese Smart Grid. *Automation of Electric Power Systems* 33(9), 1–4 (2009)
5. Chen, S., Song, S., Li, L.: Survey on Smart Grid Technology. *Power System Technology* 33(4), 1–7 (2009)
6. Zhang, J., Chen, Z.: The Impact of AMI on the Future Power System. *Automation of Electric Power Systems* 34(2), 20–23 (2010)

# An Effective Algorithm of Outlier Detection Based on Clustering

Qingsong Xia, Changzheng Xing, and Na Li

Department of Electronics and Information Engineering,

Liaoning Technical University, Huludao, China

{qinglangluck,Linna\_86}@163.com, Xcz6701@126.com

**Abstract.** This article put forward an effective algorithm to examine outlier detection based on clustering. Clustering and outlier detection methods are analyzed well by partitioning method. The results show that: it reduced time complexity of the outlier detection greatly, and with the increase of the data, this algorithm is still useful.

**Keywords:** Data mining, Clustering, K-medios, Outlier, Outlier detection.

## 1 Introduction

In data mining, there may be some data objects which do not accord with the general model of data, these data objects are called outlier, and they are different from the other parts [1].

Many algorithms of data mining try to make the effects of outlier smallest, and get rid of their influence. But in real life, because a person's noise may be another man's signal, which may lead to hide important information [2]. In other words, outlier itself may be very important, such as in the fraud detection, outlier could signal a fraud. Therefore, outlier detection and analysis is a very meaningful data mining task.

The current algorithm of outlier detection detects directly after pretreatment of data sets, needs a high level of time and space complexity [3]. A number of analyses have been designed for a database, such as clustering analysis and outlier detection etc. Clustering analysis and outlier detection do not work separately. So, the paper puts forward an algorithm of outlier detection based on clustering. This method makes outlier detection and clustering analysis combine well, and reduces outlier detection time complexity.

## 2 Relation Work

In data mining, outlier detection has four kinds: methods based on statistical analysis; method based on the distance; method based on the deviation and the method based on density.

## 2.1 Statistical Distribution-Based Outlier Detection

The statistical distribution-based approach to outlier detection assumes a distribution or probability model for the given data set (e.g., a normal or Poisson distribution) and then identifies outliers with respect to the model using a discordance test. Application of the test requires knowledge of the data set parameters (such as the assumed data distribution), knowledge of distribution parameters (such as the mean and variance), and the expected number of outliers.

“How effective is the statistical approach at outlier detection?” A major drawback is that most tests are for single attribute, yet many data mining problems require finding outliers in multidimensional space. Moreover, the statistical approach requires knowledge about parameters of the data set, such as the data distribution. However, in many cases, the data distribution may not be known. Statistical methods do not guarantee that all outliers will be found for the cases where no specific test was developed, or where the observed distribution cannot be adequately modeled with any standard distribution.

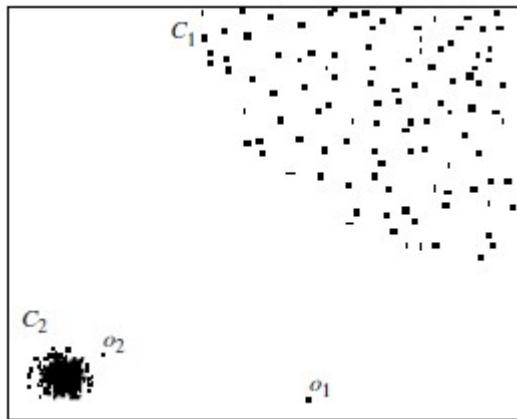
## 2.2 Distance-Based Outlier Detection

The notion of distance-based outliers was introduced to counter the main limitations imposed by statistical methods. An object,  $o$ , in a data set,  $D$ , is a distance-based (DB) outlier with parameters  $pct$  and  $dmin$ , if that is, a DB ( $pct; dmin$ )-outlier, if at least a fraction,  $pct$ , of the objects in  $D$  lie at a distance greater than  $dmin$  from  $o$ . In other words, rather than relying on statistical tests, we can think of distance-based outliers as those objects that do not have “enough” neighbors, where neighbors are defined based on distance from the given object. In comparison with statistical-based methods, distance based outlier detection generalizes the ideas behind discordance testing for various standards distributions. Distance-based outlier detection avoids the excessive computation that can be associated with fitting the observed distribution into some standard distribution and in selecting discordance tests.

## 2.3 Density-Based Local Outlier Detection

Statistical and distance-based outlier detection both depend on the overall or “global” distribution of the given set of data points,  $D$ . However, data are usually not uniformly distributed. These methods encounter difficulties when analyzing data with rather different density distributions, as illustrated in the following example.

Example necessity for density-based local outlier detection. Figure 1 shows a simple 2-D data set containing 502 objects, with two obvious clusters. Cluster  $C_1$  contains 400 objects. Cluster  $C_2$  contains 100 objects. Two additional objects,  $o_1$  and  $o_2$  are clearly outliers. However, by distance-based outlier detection. (which generalizes many notions from statistical-based outlier detection), only  $o_1$  is a reasonable( $pct, d_{min}$ )-outlier, because if  $d_{min}$  is set to be less than the minimum distance between  $o_2$  and  $C_2$ , then all 501 objects are further away from  $o_2$  than  $d_{min}$ . Thus,  $o_2$  would be considered a DB ( $pct, d_{min}$ )-outlier, but so would all of the objects in  $C_1$ ! On the other hand, if  $d_{min}$  is set to be greater than the minimum distance between  $o_2$  and  $C_2$ , then even when  $o_2$  is not regarded as an outlier, some points in  $C_1$  may still be considered outliers.



**Fig. 1.** The necessity of density-based local outlier analysis

#### 2.4 Deviation-Based Outlier Detection

Deviation-based outlier detection does not use statistical tests or distance-based measures to identify exceptional objects. Instead, it identifies outliers by examining the main characteristics of objects in a group. Objects that “deviate” from this description are considered as outliers. Hence, in this approach the term deviations are typically used to refer to outliers. In this section, we study two techniques for deviation-based outlier detection. The first sequentially compares objects in a set, while the second employs a Lapidate cube approach.

In addition, there are the method based on entropy and the method based on neural network etc.

### 3 Based on Clustering Outlier Detection Algorithm

When we analyzed a data set, because of the difference of the analysis of clustering algorithms, the outlier treatment was also different. Some algorithms in the cluster analysis did not eliminate outlier, such as dividing method and the method based on grid. And some removed the outlier as noise, for example, the method based on density method [4]. However outlier detection range was the whole data set, and was not in each cluster. If the clustering algorithm got rid of outlier as noise, it destroyed the integrity of the data set, which reduces the accuracy of the algorithm. In fact, the outlier in the clustering algorithms is eliminated because not enough "neighbors" and more is probably true outlier. Based on the above analysis, in the outlier detection process it should be first from clustering analysis and get rid of the isolation of the sites for testing. If the number of to be removed points is less than the number of outlier to test, then we test the outlier detection from each cluster.

The data set S was divided into n clusters of  $C_1, C_2, \dots, C_n$ , and got rid of m outliers ( $m >= 0$ ) of  $O_1, O_2, \dots, O_m$ , the results are  $S = \{C_1, C_2, \dots, C_n\} \cup \{O_1, O_2, \dots, O_m\}$ . The center of each cluster is  $C_j+$ , the distance between  $O_i$  and  $C_j+$  is  $d_{ij}$ .

$$d_{ij} = ((|x_{i1} - x_{j1}|^q + |x_{i2} - x_{j2}|^q + \dots + |x_{ip} - x_{jp}|^q)^{1/q})$$

P is behalf of the dimensionality. When  $q = 1$ , it shows that Manhattan distance; When  $q = 2$  said Euclidean distance. The distance from  $O_i$  to  $C_j+$  is  $d_{ij}$ , which is available in the matrix R among which the first I elements to each cluster said,  $O_i$  center distance:

$$R = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \dots & \dots & & \dots \\ d_{n1} & d_{n2} & \dots & d_{nn} \end{bmatrix}$$

$d_i = \sum_{j=1}^n d_{ij}$   $d_i$  is the sum of the first i row, and it is the distance from a leaving point to

the center, the greater the value of  $d_i$  to the center, the far distance they have.

When  $m \geq k$ , the maximum  $d_i$  before k leaving objects is to considered outliers.

When  $m < k$ , this m leaving points are to isolate points, and then again from each cluster detected in the rest of the ( $k-m$ ) an outlier.

Data object  $s_j$  in a cluster of  $C_i$  to the distance of their own cluster center  $C_j+$  is  $r_{ij}$ . Selecting ( $k-m$ ) objects from each cluster is used as a new leaving point. So the n clusters got rid of the points of  $n^*(k-m)$ . Because  $n \geq 1$ ,  $n^*(k-m) \geq k-m$ . And then from these  $n^*(k-m)$  it detected ( $k-m$ ) an outlier.

#### Algorithm 1. Clustering -Based Outlier Detection

Input: S of data sets already clustered and expected outlier number of k

Output: o of outlier set

if ( $m \geq k$ )

Choose\_max ( $d[m]$ ,  $k$ )  $\rightarrow O$ ; // Choose from an array of the maximum d before in isolated point set the number of k 0  
else

{

$d \rightarrow O$ ; // all points of the clustering should be put into outlier concentration

For ( $i=1$ ;  $i++$ ;  $i \leq n$ )

Choose\_max ( $r[i]$  [],  $(k-m)$ )  $\rightarrow d$ ; // each cluster from selected  $r[I]$  [] before the maximum ( $k-m$ ) objects in put d

Choose\_max ( $d[n^*(k-m)]$ ,  $(k-m)$ )  $\rightarrow O$ ; // select values from an array d the biggest former ( $k-m$ ) number into isolated point set of O

}

$Cout \ll O$ ; // Output isolated point set

## 4 The Algorithm Analysis and Comparison

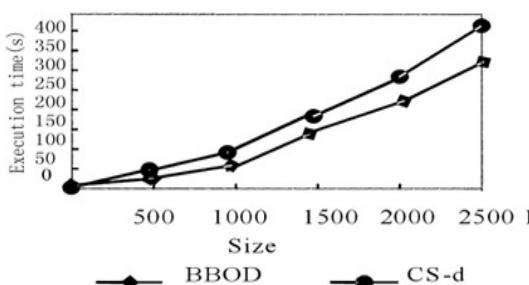
Outlier has a lot of different definitions, one of which is the biggest distance from the average distance of k. A single particle of center represents the entire object in BCOD

algorithm. If changing distance [2] of the cluster, the first is taken out far from the point of center. Because of the number of points off far smaller than the number of objects in the cluster data, so it can be neglected. Therefore, it is equal to the distance of the first  $i$  leaving points to the entire data object. Because the average distance = distance and/ $k$ , and distance and maximum also is equivalent to the largest average distance. Therefore, the outlier who is detected by BCOD algorithm meets the definition of outlier can assure the correctness of the algorithm. BCOD algorithm uses the center of  $k$ -center clustering to be as the center of heart, reduces the error which the clustering algorithm for outlier. The number of points which is getting rid of every cluster is equal to the number which will be detected. So the effect is more close to the real results to avoid more error.

In BCOD algorithm, a calculation matrix  $R$  and the complexity of the  $d_i$  are  $O(m * n)$ . When  $m >= k$ , selected from  $m$  number the largest number  $k$  the complexity of the  $O(m * \text{for } k)$ . So, when  $m >= k$ , BCOD algorithm complexity for Max ( $O(m * n)$ ,  $O(m * k)$ ). When  $m < k$ , the computing  $r[I]$  the complexity of the [] for  $O(N)$ , from each cluster selected  $r[I][\cdot]$  before the maximum ( $k-m$ ) the complexity of the object for  $O(N * N * (k-m))$ , from  $m-N * (k-the \text{ number of } m)$  before selecting ( $k-the \text{ most } m$ ) the complexity of the large for  $O(N * (k-m))$ , and 2) Because  $\text{Max}(n, k) << n$ , so when  $m < k$ , the computational complexity of BCOD algorithm for  $O(n * n * (k-m))$ .

Therefore, BCOD algorithm under the condition of the worst time complexity is  $O(n * n * (k-m))$ . The complexity of data objects and dimension, therefore, this algorithm has good expansibility.

To further validate the BCOD algorithm, we realize BCOD algorithm by vc++ 512 M in memory, hard drive for 40 G, Pentium CPU for IV 2.4 G, the operating system for Windows 2000 Server. The experiment uses the NHL data set of the literature [3], records the data sets of American hockey federation, such as the information of the athletes' personal data, the scores of every game, etc. The experimental results as shown in figure 2 below, which CS-d algorithm is, belong to the disk algorithm based on the reference [1]. Obviously, this paper puts forward algorithm, in which efficiency is superior.



**Fig. 2.** Different algorithms have different results

## 5 Conclusions

Clustering analysis and outlier detection are often together in data mining process. BCOD clustering algorithm realizes outlier detection based on clustering. In the outlier

detection process, this algorithm does not need to input parameters, so it avoids the inconvenience and error caused for the improper parameter. In addition, the algorithms are simple, intuitive, and realize the principle, and the time complexity and scalability is better than the other outlier detection algorithm proposed in the references. Therefore, BCOD algorithm has its unique advantage in the data set on outlier detection.

## References

1. Angiulli, F., Pizzuti, C.: Fast outlier detection in high dimensional space. In: Proceedings of the Sixth European Conference on the Principles of Data Mining and Knowledge Discovery, pp. 15–16 (2002)
2. Han, J., Micheline, K.: Data mining and technology, FanMing, MengXiaoFeng (transl.), pp. 223–259. Mechanical industry press, Beijing (2002)
3. Papadimitriou, S., Kitawaga, H., Gibbons, P., Faloutsos, C.: LOCI: Fast outlier detection using the local correlation integral. In: Proc. of the International Conference on Data Engineering, pp. 315–326 (2003)
4. Bay, S., Schwab Cher, M.: Mining distance-based outliers in near linear time with randomization and a simple pruning rule. In: SIGKDD 2003, Washington, DC, USA (2003)
5. Chiang, J., Yin, Z.: Unsupervised minor prototype detection using an adaptive population partitioning algorithm. Pattern Recognition 40, 3132–3145 (2007)
6. Barnett, V., Lewis, T.: Outliers in Statistical Data. John Wiley & Sons, New York (1994)
7. Zhang, Q., Couloigner, I.: A new and efficient K-medoid algorithm for spatial clustering. In: Gervasi, O., Gavrilova, M.L., Kumar, V., Laganá, A., Lee, H.P., Mun, Y., Taniar, D., Tan, C.J.K. (eds.) ICCSA 2005. LNCS, vol. 3482, pp. 181–189. Springer, Heidelberg (2005)
8. Ramswami, S., Rastogi, R., Shim, K.: Efficient Algorithm for Mining Outliers from Large Data Sets. In: Proc. ACM SIGMOD, pp. 427–438 (2000)

# Research on Multimedia Signal Acquisition Strategy Based on Compressed Sensing

Xiaohua Guo

Fair Friend Institute of Electromechanics,  
Hangzhou Vocational and Technical College  
Hangzhou 310018, China  
Xiaohua-guo@hzvtc.edu.cn

**Abstract.** With the limited resources, it is difficult for wireless sensor network to achieve a long time, continuous, high-speed acquisition of multimedia information and a real-time, reliable transmission of high-volume sampling data. This article proposed the novel data acquisition strategy based on compressed sensing theory, which can perfectly achieve the long time, real-time, reliable transmission of high-volume multimedia data in wireless sensor network. Reasonable experiments were designed to verify the effectiveness of the algorithms, and the experiment results show that: the proposed multimedia signal acquisition strategy is reasonable, practicable, and more suitable for the wireless multimedia sensor networks.

**Keywords:** wireless sensor networks, compressed sensing, random sampling, multimedia signal.

## 1 Introduction

Comparing with the traditional wireless sensor networks, wireless multimedia sensor networks introduce extensive audio, video, images and other multimedia information. As the limited bandwidth of wireless sensor networks, and the limited processing capacity, energy, storage capacity of sensor node, it is difficult to wireless multimedia sensor networks to achieve a long time, continuous, high-speed acquisition of multimedia information and a real-time, reliable transmission of high-volume multimedia sampling data. How to solve the conflict between limited resources of wireless multimedia sensor networks and high-volume multimedia data transmission is a big issue.

In order to resolve the conflict, many researchers design or modify MAC protocols and routing protocols, but the multimedia information acquisition and processing technology research is also very important issue for helping to resolve the conflict. At present, the common sensor network data compression algorithms, such as data compression method using histograms[1], data fusion method[2], and compression algorithms based on wavelet are suitable for data monitoring tasks of statistical, reasoning, not care about the middle data. These algorithms can meet the data transmission requirements of multimedia sensor networks to some extent. This article proposes the novel data acquisition strategy based on compressed sensing theory [3],

which can perfectly achieve the long time, continuous, high-speed acquisition of multimedia information and a real-time, reliable transmission of high-volume sampling data.

The detailed algorithm flow charts are presented in this paper, and appropriate experiments are designed to verify the effectiveness of the proposed algorithm. Experiments results show that the proposed multimedia signal acquisition strategy is reasonable and practicable to achieve the long time, remote, real-time, reliable transmission of high-volume acoustic signal measurement data.

## 2 Multimedia Signal Acquisition Strategy Based on Compressed Sensing Theory

In this section we will first introduce the framework of compressed sensing theory, and then propose the data acquisition algorithm based on compressed sensing theory.

### 2.1 Overview of Compressed Sensing Theory

CS theory asserts that one can recover certain signals and images from far fewer samples or measurements than traditional methods use. The overall framework of compressive sensing theory is as follows.

Considering a finite one-dimensional discrete-time signal  $x$ , which we can expand in an orthonormal basis as follows:

$$x = \psi z, \quad \|z\|_0 = k \quad (1)$$

Where:  $x$  is a finite one-dimensional discrete-time signal,  $x \in R^{n \times 1}$ ;  $\psi$  is an orthonormal basis,  $\psi \in R^{n \times n}$  [4];  $z$  is the coefficients vector of  $x$  on the orthonormal basis  $\psi$ ,  $z \in R^{n \times 1}$ ;  $n$  is the signal dimension.  $\|z\|_0$  is the L0 norm of vector  $z$ . The vector  $z$  is a sparse vector when all but a few of its entries are zero. If there are only  $k$  nonzero entries in the  $z$ , we will call the signal  $x$  a  $k$ -sparse signal on the basis  $\psi$ .

On this basis, if we can construct a measurement matrix  $\phi$ ,  $\phi \in R^{m \times n}$  [5], we can get (2),  $y \in R^{m \times 1}$ . If we want an approximation of  $x$  we need to solve the L1-norm optimization problem, which is shown in (3). The problem (3) is a convex optimization problem, which can be easily simplified to the linear programming problem, and then we can use interior point method, gradient projection method, second-order cone programme method, and matching pursuit method to solve the problem [6].

Suppose  $\hat{z}$  is the solution to the optimization problem, then we get (4). The reconstruction error can be calculated by (5).

$$y = \phi x \quad (2)$$

$$\hat{z} = \arg \min \|z\|_1 \quad s.t. \quad y = k x z \quad (3)$$

$$\hat{x} = \psi \hat{z} \quad (4)$$

$$e = \frac{\|\hat{x} - x\|_2}{\|x\|_2} \quad (5)$$

Where:  $\hat{z}$  is the solution to the optimization problem (3);  $\hat{x}$  is the approximation of  $x$ ;  $\|z\|_1$  is the L1-norm of  $z$ ;  $\|z\|_2$  is the L2-norm of  $z$ .

## 2.2 Random Sampling Strategy

Effective way to solve the problems [7] of the existing data acquisition algorithm is to randomly sample the multimedia signal[8]. According to compressed sensing theory, when  $x$  is the sampled data, (2) can be viewed as the compression of sampling data, when  $x$  is multimedia signal, (2) can be viewed as random sampling of the acoustic signal. Random sampling process can be regarded as the combination of signal sampling and data compression.

The key to random sampling is how to get reasonable random sampling time sequence. Common method is simple random sampling [9], using random number generator to randomly generate sampling time. The main drawback of this method is that the sampling interval is hard to control, it is often too short or too long, so the result is that the hardware is difficult to meet the requirements of high frequency sampling or a certain signal sampling frequency is too low. In response to these shortcomings mentioned above, we propose the following improved random sampling method.

First, according to the signal characteristics, storage capacity, reconstruction errors to determine the appropriate length of sampling time window  $T_{sw}$ , and define  $\alpha$  sampling time  $t_i$  for each sampling window, and then we can introduce the following additive sampling process.

$$t_i = t_{i-1} + \tau_i \quad t_0 = 0 \quad (6)$$

Where:  $t_i$  is the i-th sample moment,  $\tau_i$  is the sample interval.

$$\begin{aligned} \tau_i &= \tau_{sample} + \tau_{jitter} \\ &= N\left(\frac{T_{sw}}{\alpha}, \mu^2 \frac{T_{sw}^2}{\alpha^2}\right) + s \times N(0,1) \end{aligned} \quad (7)$$

Where:  $T_{sw}$  is the sampling time window,  $\alpha$  is the sampling frequency,  $\mu$  is the convergence factor,  $s$  is the sampling time change factor.  $N(a,b)$  is the Gaussian random variable[10],  $a$  is the mean,  $b$  is the variance.

The random sampling method can be achieved by software programming. sensor nodes first set the random number generator, random number seeds, sampling time window, sampling frequency and other parameters, and then use (7) to calculate random number, use (6) to calculate the random sampling time which is used for setting and starting the timer interrupt; when the timer interrupt arrives, the sound signal sampling components will be started to sample the acoustic signal, and then repeat the steps of signal sampling mentioned above; at last the sampled data is packaged and sent to the remote computer.

When the remote computer receives the sampled data packet, it calls the pre-stored, random number generator, random number seeds, sampling time window, sampling frequency and other parameters to construct a suitable measurement matrix[11], and reads the corresponding number of random sampling data, and then uses (3), (4) to reconstruct the sound signals, finally, analysis and processes the reconstructed sound signal.

### 2.3 Discussion

According to sparse matrix dimension, measurement matrix dimension and signal reconstruction accuracy, the values of  $T_{sw}$  and  $\alpha$  can be obtained by calculating (8), and the corresponding average sampling period can be determined by (9).

$$T_{sw} = \frac{n}{f_{ss}} \quad \alpha = m \quad (8)$$

$$f_{rs} = \frac{\alpha}{T_{sw}} \quad (9)$$

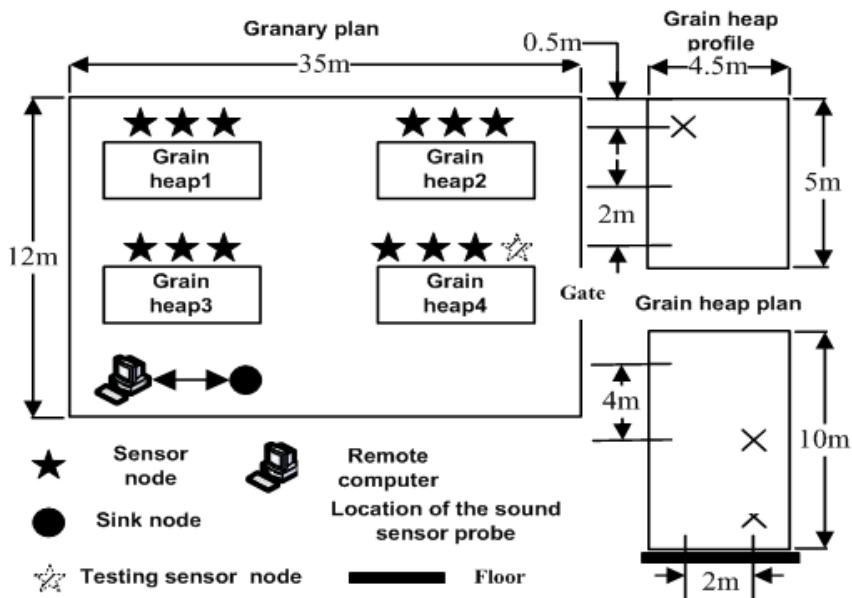
Where:  $f_{ss}$  is the frequency of Shannon sampling,  $f_{rs}$  is the frequency of random sampling.

When  $\tau_i$  is the random number defined by (7), the probability density function of the sampling time sequence defined by (6) can converge to evenly distribution the same with simple random sampling. It is a effective way to avoid the sampling interval is too large or too small, and the sequence of sampling time is more clear.

## 3 Experimental Section

### 3.1 Experimental Device

We chose one grain storage in Hangzhou to do simulation experiment. The algorithms were introduced to the stored grain pests acoustic signal acquisition system for collecting *Tribolium castaneum* adults crawling sound. The distribution of sensor nodes, sink node and sound sensor probes is shown in Fig.1.



**Fig. 1.** Simulated experimental system

The address of each sensor node is considered as the random number seed. We make all the sensor nodes sending complete sampled data to remote computer.  $\psi$  is Fourier transform matrix,  $\phi$  is Gaussian random matrix, and we use the interior point method to solve (3).  $p=12$ ,  $\mu = 0.25$ ,  $s = 2$  ms,  $f_{ss} = 8$  KHz. We determine  $n$ ,  $m$ ,  $f_{rs}$ ,  $T_{sw}$ ,  $\alpha$  and other parameters by experiments.

### 3.2 Experimental Program

We design 3 experiments to verify and compare the performance of the proposed algorithm and the existing data compression algorithm[1]. First, we use a recorder to collect a certain length sound signal, and then process them on the computer to determine those parameters: the maximum reconstruction error,  $n$ ,  $m$ ,  $f_{rs}$ ,  $T_{sw}$  and  $\alpha$ .

We set the number of acquisition channel to be 1 (only 1 channel of sensor probe is sampled), open all the nodes, use a recorder to collect 32 sound signals (each lasted about 10s), and then respectively compare the recorded signals to the signals reconstructed by the computer. We can get the reconstruction errors from the comparison and the packet loss rate by calculating (10).

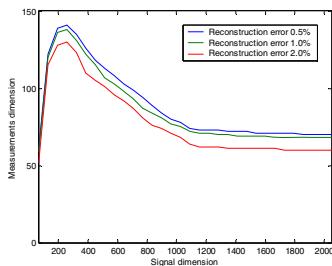
$$pl = \frac{ps - pr}{ps} \quad (10)$$

Where:  $ps$  is the number of data packets sent by sensor node,  $pr$  is the number of packets received by remote computer.

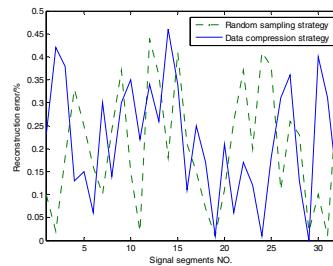
Based on experiment 2, change the number of sampling channel of each sensor node(increase 1 channel in every 10min), analyze the changes of packet loss rate with the load changes. According to the analysis results, we change the parameters of test device, and recalculate the packet loss rate and reconstruction error of testing node using the same methods with experiment 2.

### 3.3 Experimental Results

The main results are shown in Fig.2~3.



**Fig. 2.** Relationship between acoustic signal dimension, measured value dimension and signal recovery error



**Fig. 3.** Recovery error of 32 segments acoustic signal

Fig.2 shows that when  $e$  is constant, the larger the  $n$ , the smaller the  $m$ ; when  $n$  is constant,  $e$  will increase with the decrease in  $m$ . Considering the resource consumption, we set the maximum reconstruction error to be 0.5%,  $n=1024$ ,  $m=75$ ,  $T_{sw}=128\text{ms}$ ,  $\alpha=75$ ,  $f_{rs}=586\text{Hz}$ .

Fig.3 shows that, when the experimental device runs data compression strategy, the maximum error between original signal and reconstructed signal is 0.46%; when running random sampling strategy, the maximum error between original signal and reconstructed signal is 0.44%. This shows that: the two acquisition strategies both can achieve the wireless, remote, distributed acquisition of sound signals.

## 4 Conclusions

Based on the compressed sensing theory, a novel multimedia signal acquisition strategy is proposed in this article. Experiments results show that the proposed multimedia signal acquisition strategy can effectively solve the conflict between high-volume data transmission and limited network resource. The proposed algorithm based on random sampling strategy with those advantages of small amount of calculation, small footprint, low power consumption is more suitable for the wireless multimedia sensor networks.

## References

- [1] Grimberg, R., Savin, A.: Fuzzy inference system used for a quantitative evaluation of the material discontinuities detected by eddy current sensors. *Sensors and Actuators* 81(3), 248–250 (2000)
- [2] Odeberg, H.: Distance measure for sensor opinions. *Measurement Science and Technology* 4(8), 808–815 (1993)
- [3] Donoho, D.: Compressed sensing. *IEEE Trans. Information Theory* 52, 1289–1306 (2006)
- [4] Donoho, D., Tsaig, Y.: Extensions of compressed sensing. *Signal Processing* 86, 533–548 (2006)
- [5] Candes, E., Romberg, J., Tao, T.: Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Information Theory* 52(4), 489–509 (2006)
- [6] Candes, E., Tao, T.: Decoding by linear programming. *IEEE Trans. Information Theory* 51, 4203–4215 (2005)
- [7] Gastpar, M., Vetterli, M.: Power, spatio-temporal bandwidth, and distortion in large sensor networks. *IEEE Journal Select. Areas Communication* 23, 745–754 (2005)
- [8] Candes, E., Tao, T.: Near optimal signal recovery from random projections: universal encoding strategies. *IEEE Trans. Information Theory* 52, 5406–5425 (2006)
- [9] Boyle, F., Haupt, J., Fudge, G.: Detecting signal structure from randomly sampled data. In: Proceeding of 2007 IEEE Workshop on Statistical Signal Processing, Madison, Wisconsin, USA, pp. 326–330 (2007)
- [10] Zainul, C., Young, H.K., Sadaf, Z.: Energy efficient sampling for event detection in wireless sensor network. In: 2009 International Symposium on Low Power Electronics and Design, San Francisco, California, USA, pp. 587–593 (2009)
- [11] Davenport, M., Duarte, M., Wakin, M.: The smashed filter for compressive classification and target recognition. In: Proceeding of 2007 Computational Imaging V at SPIE Electronic Imaging, San Jose, California, USA, pp. 326–330 (2007)
- [12] He, T., Stankovic, J.A., Marley, M.: Feedback control-based dynamic resource management in distributed real-time systems. *Journal of Systems and Software* 80, 997–1004 (2007)

# PFSA: A Novel Fish Swarm Algorithm

Zushun Wu<sup>1</sup>, Zhangji Zhao<sup>2</sup>, Sisi Jiang<sup>1</sup>, and Xuechi Zhang<sup>1</sup>

<sup>1</sup> College of Information Science and Technology  
Beijing Normal University, 100875, Beijing, China

<sup>2</sup> Department of Physics, School of Science  
Harbin Institute of Technology, 150001, Harbin, China  
{wuzs100, zhaozhangchris, jiangsisinol, z\_xc08}@126.com

**Abstract.** A novel intelligent optimization algorithm called as particle fish swarm algorithm (PFSA) is presented in this paper. PFSA is inspired from artificial fish swarm algorithm (AFSA) and particle swarm optimization (PSO), and it can overcome the weakness of AFSA to quickly and precisely search out the optimum solution in optimization problems. In order to study the performance of PFSA, four well-known benchmark functions are optimized by PFSA and AFSA in the simulation experiments, the results show that the performance of PFSA is better than AFSA. In other word, PFSA has good performances (e.g. fast convergence speed and nice precision) to achieve the global best solution in the optimization problems.

**Keywords:** particle fish swarm algorithm, artificial fish swarm algorithm, function optimization.

## 1 Introduction

With the development of our society, more and more complex practical optimization problems appear in academic and industrial fields, so how to solve those complex problems effectively and economically has been an important issue. In the sphere of optimization, a number of classical intelligent optimization algorithms are proposed, such as particle swarm optimization (PSO) [1, 2] and artificial fish swarm algorithm (AFSA) [3]. PSO proposed in 1995 is a random search and optimization algorithm, and it is inspired by a flock behavior of bird. AFSA is inspired by the collective behavior of fish such as preying, swarming, following and random behaviors. These classical algorithms can be applied to deal with a lot of complex real-world optimization problems. For instance, PSO can be used to design water supply system [4] or improved to solve constrained optimization problems [5] or other fields; AFSA can be applied to forecast stock index [6] or other fields [7-10].

The classical optimization algorithms like PSO and AFSA have gained considerable acceptance, however, there still exists certain disadvantages in these algorithms. For example, AFSA has the worse precision and slower convergence speed. It is impossible that an algorithm is suitable to solve all problems. Accordingly, for coping with more complex problems, it is necessary to improve the

current algorithms or propose a new algorithm. In this paper, a novel intelligent algorithm called as PFSA is presented to overcome the weakness of AFSA.

PFSA is inspired from AFSA and PSO. Like AFSA, the four behaviors of fish swarm (i.e. preying, swarming, following and random behaviors) are also included in PFSA, and in the meantime the idea of PSO is introduced into PFSA, such that, PFSA can search out the global optimum solution effectively. Additionally, in order to profoundly analyze the performance of PFSA, four well-known benchmark functions are introduced into our simulation experiments.

## 2 Artificial Fish Swarm Algorithm

Artificial fish swarm algorithm (AFSA) which is an artificial intelligent algorithm is proposed in 2003 [3], it simulates the four behaviors of fish school, such as preying behavior, swarming behavior, following behavior and random behavior. Like other optimization algorithms, AFSA also can be applied to solve plenty of optimization problems [6-10].

In the  $D$ -dimensional searching space, suppose there are  $N$  fishes in the colony, the current position of artificial fish  $i$  is denoted as  $X_i = (x_1^i, x_2^i, \dots, x_D^i)$ , then the food consistence of  $X_i$  can be represented as  $Y_i = f(X_i)$ , where  $i \in \{1, 2, \dots, N\}$  and  $f$  is the objective function. Moreover, the distance between artificial fish  $i$  and  $j$  is expressed as  $d_{ij} = \|X_j - X_i\|$ , where  $j \in \{1, 2, \dots, N\}$ . Consequently, in AFSA, the four behaviors of fish swarm are as follows.

### A. Preying behavior

Suppose the visual field and position of artificial fish  $i$  can be denoted as *Visual* and  $X_i$  respectively, and the artificial fish  $i$  randomly selects a new position  $X_j$  ( $d_{ij} < \text{Visual}$ ), then if  $Y_i > Y_j$  in the minimum problem (we only consider the minimum problem in this paper, there will be opposite condition in the maximum problem), the artificial fish  $i$  moves a step following the expression (1). Otherwise it randomly selects a new  $X_j$  again, if the condition  $Y_i > Y_j$  is still unsatisfied when the try-number excesses a preset value, then the preying behavior will be quitted.

### B. Swarming behavior

We assume that the companion number of the artificial fish  $i$  in its visual field is  $N_f$ , the central position of those fishes is  $X_c$  and  $\delta$  is a crowded degree factor. If  $Y_c N_f < \delta Y_i$ , which means that the position  $X_c$  is better than  $X_i$  and it is not crowded in the position  $X_c$ , the artificial fish  $i$  moves a step following the expression (1), otherwise, gives up the swarming behavior.

### C. Following behavior

In the visual field of the artificial fish  $i$ , we suppose the companion whose current state is  $X_f$  has the best  $Y_f$ . If  $Y_f N_f < \delta Y_i$ , which means it is not crowded in the  $X_f$ , the artificial fish moves a step following the expression (1), otherwise, the following behavior will not be performed.

#### D. Random behavior

An artificial fish randomly chooses a new position in its visual field, and swims following the expression (2).

$$X_i(t+1) = X_i(t) + Step \times Rand_1 \times \frac{X_v - X_i(t)}{\|X_v - X_i(t)\|} \quad (1)$$

$$X_i(t+1) = X_i(t) + Rand_2 \times Visual \quad (2)$$

where  $t$  is the number of iteration,  $Step$  is a moving step length,  $Rand_1$  is a random parameter within the interval  $[0, 1]$ ,  $Rand_2$  is a D-dimensional vector whose elements are random numbers within the interval  $[-1, 1]$ , and  $Visual$  is the visual field. Especially,  $X_v = X_j$  for the preying behavior,  $X_v = X_c$  for the swarming behavior,  $X_v = X_f$  for the following behavior.

The selection strategy of the above behaviors is that: each artificial fish judges whether preying, swarming and following behaviors can be performed. If at least one of the three behaviors can be carried out, the behavior which can let the artificial fish move to a better position will be chosen; if the artificial fish cannot perform the three behaviors, the random behavior will be selected to be its behavior.

According to the discussions in literature [3, 7, 8], there are certain disadvantages in artificial fish swarm algorithm (AFSA), such as the slower convergence speed and worse precision. On the one hand, the moving step length generally is a small fixed value in AFSA, so the convergence speed of AFSA is slower. More importantly, the fishes are impossible to be too closer to the best solution because of the fixed step. On the other hand, the crowded degree judgment of AFSA is  $Y_v N_f < \delta Y_i$  (for minimum problem), and it can help the fish school to escape from the local optimal solution. However, the judgment is unreasonable because the fishes always feel crowded when they are too close to the optimal solution (i.e.  $Y_i \approx Y_v$ ), that is to say, the fishes are impossible to move to the optimum solution with a good precision.

In order to improve the performance of AFSA, the problems discussed above should be well solved. For this purpose, a novel intelligent optimization algorithm, which has a fast convergence speed and good precision, is presented and discussed in the following sections.

### 3 Particle Fish Swarm Algorithm

Inspired from AFSA and PSO, we present a novel intelligent optimization algorithm called as particle fish swarm algorithm (PFSA). In our PFSA, the idea of social part of PSO is introduced, and the moving step length and crowded degree judgment of PFSA are different from those of AFSA.

In fact, PFSA is similar to AFSA to some extent. The preying, swarming, following and random behaviors are also included in PFSA and the selection strategy of PFSA is the same as that of AFSA. The differences between PFSA and AFSA are as follows.

### A. Expressions

The expressions of PFSA which are different from those of AFSA are shown as expression (3) and (4), where the random behavior is denoted by expression (4) and other behaviors are denoted as expression (3).

$$X_i(t+1) = X_i(t) + c_1 r_1 (X_v - X_i(t)) + c_2 r_2 (G_{best} - X_i(t)) \quad (3)$$

$$X_i(t+1) = X_i(t) + r_3 \times Step \quad (4)$$

where  $t$  is the number of iteration,  $c_1$  and  $c_2$  which are positive constants are called as cognitive and social parameters respectively,  $r_1$  and  $r_2$  are random parameters within the interval  $[0, 1]$ ,  $G_{best}$  is the global optimum solution at the  $t$ th iteration.  $r_3$  is a D-dimensional vector whose elements are random numbers within the interval  $[-1, 1]$ , and  $Step$  is the moving step length in random behavior. Like AFSA,  $X_v = X_j$  for the preying behavior,  $X_v = X_c$  for the swarming behavior,  $X_v = X_f$  for the following behavior.

---

```
//PFSA
Initialize all basic parameters of PFSA;
for i=1 to the swarm size N do
    Initialize the position  $X_i$  of fish i;
end for
    Get the current global optimum solution  $G_{best}$ ;
while (t < maximum iteration times) do
    for i=1 to the swarm size N do
        for try-number =1 to the maximum try-number do
            Randomly find a better position  $X_j$  in the visual of  $X_i$ ;
        end for
        for k=1 to the swarm size N do
            Find the companions of  $X_i$  in its visual;
            Evaluate the positions of the companions;
        end for
        Get the center  $X_c$  of those companions;
        Find the best fish position  $X_f$  in  $X_i$ 's visual;
        if (the condition of random behavior is satisfied) then
             $X_i(t+1) = X_i(t) + r_3 \times Step$  ;
        else
            Select the best position denoted as  $X_v$  from  $X_j$ ,  $X_c$  and  $X_f$  ;
             $X_i(t+1) = X_i(t) + c_1 r_1 (X_v - X_i(t)) + c_2 r_2 (G_{best} - X_i(t))$ ;
        end if
    end for
    Get the current global optimum solution  $G_{best}$ ;
end while
```

---

**Fig. 1.** The pseudo-code of PFSA

### B. Crowed degree judgment

In a fish school, suppose there are  $N$  fishes in the searching space, the purpose found by fish  $i$  is  $X_v$ , the distance between  $X_i$  and  $X_v$  is  $d_{iv}$ . The fish  $i$  computes all

distances between all positions of companions in its visual and the purpose  $X_v$ , then ranges those distances by size from short to long and gets the sort order of  $d_{iv}$  denoted as  $s_i$ . If  $s_i < \delta N$ , which means the fish  $i$  is closer to the purpose and competitive to find food, then the purpose position  $X_v$  is not crowded for fish  $i$ , where  $\delta$  is a crowded degree factor. Otherwise,  $X_v$  is crowded for fish  $i$ .

According to the above descriptions, the pseudo-code of PFSA can be shown as Figure 1. In our PFSA, the idea of the social part of PSO (i.e. the global optimum solution  $G_{best}$ ) is introduced, and the moving step length is no longer fixed but can be adjusted according to the purpose position, so the convergence speed could be enhanced. Additionally, the crowded degree judgment in PFSA is better than that in AFSA, because the judgment in PFSA can make the fishes who are closer to the purpose position continue moving to the purpose when  $Y_i \approx Y_v$ , such that, the precision can be improved. In a word, our PFSA can let the fish swarm find out the best value more quickly and precisely.

## 4 Simulation Experiment

In order to evaluate the performance of PFSA, four well known benchmark functions are used in our simulation experiment. The four functions are as follows.

$$1) \text{ Rosenbrock Function: } F_1 = 100(x_1^2 - x_2)^2 + (1 - x_1)^2$$

$$\text{Range} = [-2.048, 2.048]^2, \quad F_1(\min) = 0.$$

$$2) \text{ Sphere function: } F_2 = \sum_{i=1}^2 x_i^2$$

$$\text{Range} = [-10, 10]^2, \quad F_2(\min) = 0.$$

$$3) \text{ Easom function: } F_3 = -\cos(x_1)\cos(x_2)\exp(-(x_1 - \pi)^2 - (x_2 - \pi)^2)$$

$$\text{Range} = [-10, 10]^2, \quad F_3(\min) = -1.$$

$$4) \text{ Hump function: } F_4 = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4$$

$$\text{Range} = [-5, 5]^2, \quad F_4(\min) = -1.0316.$$

In the simulation, the above four benchmark functions are optimized by our PFSA and AFSA in MATLAB. The maximum iteration times is 100, the size of swarm is

**Table 1.** The initial parameters of PFSA and AFSA

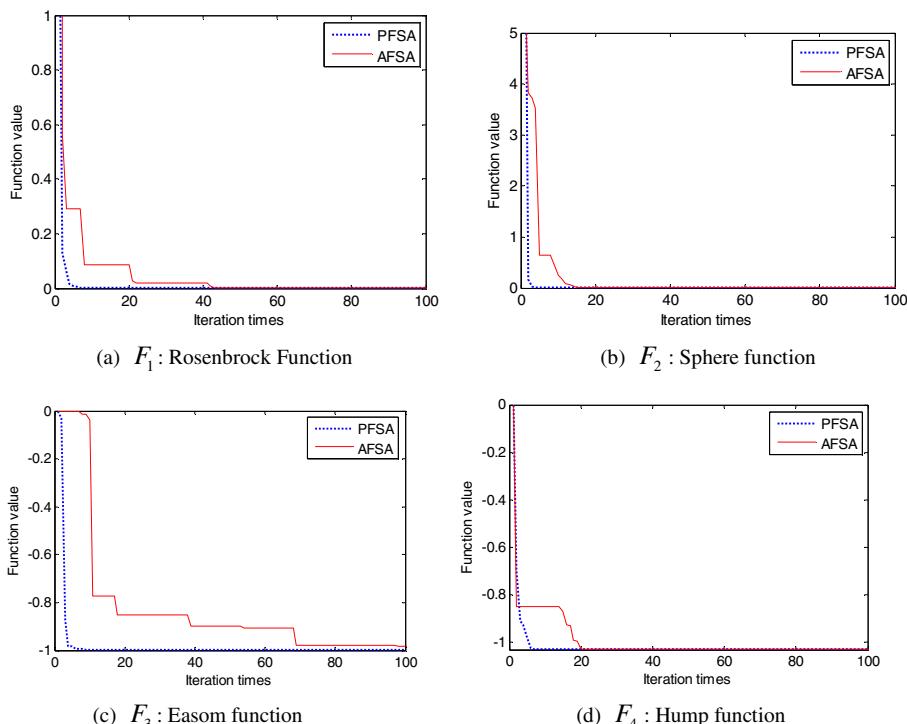
Function	PFSA		AFSA	
	Visual	Step <sub>max</sub>	Visual	Step
$F_1$	0.5	0.5	0.5	0.1
$F_2$	2.5	2.5	2.5	0.3
$F_3$	2.5	2.5	2.5	0.3
$F_4$	1	1	1	0.1

$N=20$ , the crowed degree factor is  $\delta = 0.618$ , the cognitive parameter  $c_1$  and social parameter  $c_2$  both are 2, and the try-number is 20. The step length  $Step$  in PFSA is denoted as  $Step_{max}(1-t / max\_iteration)$ . According to different cases, the other initial parameters of each algorithm are as shown in Table 1.

PFSA and AFSA are executed 20 times for each optimization function, the average experiment results and the optimization curves of each function are shown in Table 2 and Figure 2, respectively.

**Table 2.** The simulation experiment results

Function	Optimum value		Precision	
	PFSA	AFSA	PFSA	AFSA
$F_1$	0.0000	0.0000	5.3764E-011	8.1281E-006
$F_2$	0.0000	0.0000	4.4537E-012	2.6693E-006
$F_3$	1.0000	-0.9832	4.6112E-006	0.0168
$F_4$	1.0316	-1.0296	2.8409E-005	0.0020



**Fig. 2.** The optimization curves of benchmark functions

As the simulation experiment results shown in Table 2 and Figure 2, we can have that the optimization precision and convergence speed of our PFSA are better than those of AFSA. This is to say, PFSA has a better performance to find out the global optimum solution in the optimization problems.

## 5 Conclusion

Inspired from artificial fish swarm algorithm (AFSA) and particle swarm optimization (PSO), we propose particle fish swarm algorithm (PFSA) in this paper. Based on four behaviors of fish school (i.e. preying, searching, swarming and random behaviors) and the idea of PSO, our PFSA can find out the global optimum solution quickly and precisely. Additionally, in order to investigate the performance of PFSA, four well-known benchmark functions are introduced into the experiments, and the simulation results show that the optimization precision and convergence speed of PFSA are definitely better than those of AFSA.

## References

- [1] Eberhart, R., Kennedy, J.: A new optimizer using particle swarm theory. In: Proceedings of the Sixth International Symposium on Micro Machine and Human Science (1995)
- [2] Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proceedings of IEEE International Conference on Neural Networks, pp. 1942–1948 (1995)
- [3] Li, X.-L.: A new intelligent optimization method-artificial fish school algorithm. Ph.D. Thesis, Zhejiang University (2003) (in Chinese)
- [4] Montalvo, I., Izquierdo, J., Pérez, R., Tung, M.-M.: Particle Swarm Optimization applied to the design of water supply systems. Computers and Mathematics with Applications 56(3), 769–776 (2008)
- [5] Sun, C.-L., Zeng, J.-C., Pan, J.-S.: An improved vector particle swarm optimization for constrained optimization problems. Information Sciences, 1153–1163 (2011)
- [6] Shen, W., Guo, X.-P., Wu, C., Wu, D.-S.: Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm. Knowledge-Based Systems, 378–385 (2011)
- [7] Ma, H., Wang, Y.-J.: An Artificial Fish Swarm Algorithm Based on Chaos Search. In: Proceedings of the Fifth International Conference on Natural Computation, pp. 118–121 (2009)
- [8] Jiang, M.-Y., Yuan, D.-F., Cheng, Y.-M.: Improved Artificial Fish Swarm Algorithm. In: Fifth International Conference on Natural Computation, pp. 281–285 (2009)
- [9] Wang, J.-P., Hu, M.-J.: A solution for TSP based on Artificial Fish Algorithm. In: Proceedings of International Conference on Computational Intelligence and Natural Computing, pp. 26–29 (2009)
- [10] Liu, T., Liu, A.-L., Hou, Y.-B., Chang, X.-T.: Feature optimization Based on Artificial Fish-swarm Algorithm in Intrusion Detections. In: Proceedings of International Conference on Networks Security, Wireless Communications and Trusted Computing, pp. 542–545 (2009)

# **Research on Platform Integration of Equipment Support Simulation Training Systems and Integration Degree**

Yunfeng Lian, Yu Lu, LiYun Chen, and Yi Ma

Six Department,  
Shijiazhuang Mechanical Engineering College,  
Shijiazhuang, China  
lianyf100@yahoo.com.cn

**Abstract.** Due to equipment support simulation training systems are geographical distributing, independently running, and lacking association with each other, which can not meet the demand of the equipment support capability rising. It puts forward platform integration to make the equipment supporting simulation training systems distributed to an organic whole. It also designs the function structure of integrated platform and research the integration degree, which can meet the need of carrying out equipment support training systematic and organized, and meet the need of training of person who knows both of the command and technology.

**Keywords:** equipment support, platform integration, integration degree.

## **1 Introduction**

The use of equipment support simulation training system can improves the training efficiency, standardizes the training order and trained personnel, which play an important role in the support capabilities formation.

The development of scientific and technological and the accelerated pace of fighting demand a higher capacity of equipment support. The training model of relying on single platform and support the type equipment can not meet the demand. This will require integrate the existing equipment support simulation training system and enhance the training function of the system, to meet the equipment support capacity requirements for diverse tasks.

## **2 Platform Integration**

Carrying platform integration to the equipment support training simulation systems is to build an integrated platform, which is a virtual platform. It can complement all of the component systems function, which can also have the function bigger than the sum of the component systems [1]. The integrated platform function structure can be shown as Fig 1.

Platform function (command and control, equipment maintenance, material support, etc.)		Function layer	
Application conversation management		Conversation layer	
Application conversation model			
Query route	Query disassemble	Information layer	
Access control (user authorization)			
Information organize			
Communication mechanism	Communication language and protocol	Communication layer	
Information format	Information content		

**Fig. 1.** Function structure of integrated platform

### (1) Communication layer

The bottom layer of an integrated platform functional structure is the communication layer, which allows systems to be message-based communication and the mutual understanding of the meaning of the message. Communication layer provides a complete messaging specification for communication between the system and deal with the messages sent or received based on the specifications. Message specification includes communication mechanism, communication language and protocol, general content of communications and information format shared by all nodes [2].

### (2) Information layer

Information layer is responsible for the organization of information and controls the use of information. Information layer organizes the system's private information and sharing information reasonable. It can ensure the security of system's private information and carry access authorization for the user (group) who wants to access the sharing information and the exchange information. It can prevent the illegal operation to information. Information layer has efficient processing capabilities of distributed information query. It can identify local queries and distributed queries and deal with them separately [3]. It can also decompose complex distributed queries into several sub-queries that easy to be executed.

### (3) Conversation layer

Conversation layer controls systems interactions. It can analyze the interaction between the various systems based on business process needs and define the various system applications conversation model. Conversation layer provides a comprehensive conversation management and create an instance of the conversation model dynamically. It can make the conversation activate, the conversation hang, conversation restore. It can achieve the correct interactions and consistent interactions between the systems by conversation's reasoning and decision-making in accordance with predefined rules and controlling the operation of the conversation instance.

#### (4) Function layer

The top layer of an integrated platform functional structure is the function layer. It defines the various functions of the integrated platform and determines the goal of the integrated platform to achieve.

#### (5) Relationship among the layers

In the function structure of an integrated platform, the function layer defines the functional requirements and objectives, conversation layer is responsible for completion target by the interaction between the systems, information layer provides interactive content and motivation, communication layer provides the interactive message specification [4]. All levels are closely related. Low layer is the foundation to achieve the high layer function and high layer function is the goal to construct low layer.

## 3 Measure of Integration

### 3.1 Average Path Length

The distance between two nodes  $i$  and  $j$  in the network is defined as the shortest path edges of connecting the two nodes. The distance between any two nodes in the network is called the maximum network diameter (diameter), denoted by  $D$ , the

$$D = \max_{i,j} d_{ij} \quad (1)$$

The network average path length is defined as the distance between any two nodes on average, which can be described as the following.

$$\bar{L} = \frac{\sum_{i>j} d_{ij}}{S} \quad (2)$$

$S$  is the number of connected nodes in the network,

$$S = \frac{n(n-1)}{2} \quad (3)$$

$n$  is the number of nodes in the network. Substituting the above equation, the network average path length can be expressed as the following.

$$\bar{L} = \frac{2\sum_{i>j} d_{ij}}{n(n-1)} \quad (4)$$

The network average path length is also known as network characteristic path length. Although many complex networks have large number of nodes, but the network average path length is surprisingly small. Specifically, in a connected network, for the determination of network nodes average degree, the rate increase of average path length is proportional to the logarithm with network size  $N$ .

### 3.2 Degree and Degree Distribution

Degree is a simple and important concept property of a single node. Node degree is defined as the node number that connected to the node. General sense, the greater degree of a node means that the node is more important in the network. The average degree of all nodes is called the average degree of the network.

$$\bar{k} = \frac{\sum_{i=1}^n k_i}{n} \quad (5)$$

The distribution of node degree in the network can be described by the distribution function  $N_k$ , which is the number of nodes with degree  $k$ .

$$p_k = \frac{N_k}{n} \quad (6)$$

### 3.3 Clustering Coefficient

Clustering coefficient, also known as network density, is used to describe the network aggregation node, that is how close the network. Assume a node is connected with  $k_i$  nodes. If the  $k_i$  nodes are connected with each other, the edges number can be described as

$$S_i = \frac{k_i(k_i - 1)}{2} \quad (7)$$

If there are  $E_i$  edges between the  $k_i$  nodes, the clustering coefficient can be described as the ratio of  $E_i$  and  $S_i$ .

$$C_i = \frac{E_i}{S_i} = \frac{2E_i}{k_i(k_i - 1)} \quad (8)$$

Network clustering coefficient is the average of all nodes clustering coefficient.

$$\bar{C} = \frac{\sum_{i=1}^n C_i}{n} = \frac{\sum_{i=1}^n \frac{2E_i}{k_i(k_i - 1)}}{n} \quad (9)$$

$C = 0$  if and only if all nodes are isolated nodes, that is no connecting edge;  $C = 1$  if and only if the network is globally coupled, that is, any two nodes in the network are directly connected.

### 3.4 Platform Integration Degree

Given the average degree of network integrated platform, the gain of the entire network can be expressed as follows:

$$I(\bar{k}) = p(\bar{k}) - c(\bar{k}) \quad (10)$$

$I(\bar{k})$ , which describes the network connection Gains of the integrated platform.

$p(\bar{k})$ , which describes the network connection benefits of the integrated platform.

$c(\bar{k})$ , which describes the network connection costs of the integrated platform.

Gains are decided both by benefits and costs. How to improve the benefits and reduce the costs is the key of enhancing the gain.

### 3.5 The Benefits of Information Network Connection

When the network average path length is gradually reduced, the network connection benefits also increased, the network will become the biggest gain 1. So you can set the following function to express the benefits

$$f(\bar{L}) = \frac{1}{\alpha \bar{L} + 1} \quad (11)$$

The connection benefits factor .According to small-world network model, the relationship of the connection benefits and the average degree of network connections can be expressed as following.

$$f(\bar{k}) = \frac{1}{\alpha' \frac{\ln(\beta \bar{k})}{\bar{k}^2} + 1} \quad (12)$$

### 3.6 The Cost of Network Connection

Network average degree describes the average connection degree of the network nodes. As the network average degree increases, the nodes will be gradually connected over redundant information which led to information overload, Therefore, the network connection costs will increase and reduce the gain of the entire network. Therefore, the cost function can be set to the following:

$$c(\bar{k}) = 1 - e^{-v\bar{k}} \quad (13)$$

One, v indicates that the connection cost factor.

### 3.7 Network Connection Gain Discussion

According to the above formula, it can get the network average degree to make the network get the max gains

$$I(\bar{k}) = \frac{1}{\alpha' \frac{\ln(\beta \bar{k})}{\bar{k}^2} + 1} - (1 - e^{-v\bar{k}}) \quad (14)$$

The average degree can be getting to make the network obtain the most gains.

$$\frac{d(I(\bar{k}))}{d(\bar{k})} = 0 \mid \bar{k}^* = \bar{k} \quad (15)$$

If not carrying the platform integration to the equipment support simulation training systems, there is no connection benefits and no connection costs. Therefore, the integrated platform network performance is 0, which can not achieve the purpose of enhancing the training capacity of the system. For a lot of the average degree of network integrated platform, the connection benefits is saturation, but it also has a large connection costs, the integrated platform of network performance is very low. So it should be controlled to the simulation training system integrated platform degree to obtain optimal network connectivity.

## 4 The References Section

Carrying platform integration to equipment support simulation training system can make the equipment support simulation training system which is geographically dispersed, structurally heterogeneous, functionally independent, business closed to an organic system, which can change the existing system run independently, incompatible operating mode, and carry out training system and Formed.

## References

1. Fan, Y.S., Wu, C.: MACIP: solution for CIMS implementation in manufacturing enterprises. In: Proc. IEEE International Conference on Factory Automation & Emerging Technology, Los Angles, USA, pp. 1–6 (September 1997)
2. Xue, S.X., Fan, Y.S.: Global information system in manufacturing enterprise CIMS integration platform. In: Proc. IEEE Int. Conf. on Systems, Man, and Cybernetics, Beijing, pp. 1404–1407 (October 1996)
3. Wei, S., Wu, C., Fan, Y.S.: MACIP: an open structured supporting platform for CIMS. In: Proc. IEEE Int. Conf. on Systems, Man, and Cybernetics, Beijing, pp. 893–896 (October 1996)
4. Fan, Y.-S., Zhao, B.: Whole Enterprise Settle Solution Oriented Integrated Modeling and Implementing. Computer Integrated Manufacturing Systems 8(11), 841–845 (2002)
5. Foster, I., Kesselman, C., Steven, S.: The anatomy of the grid: enabling scalable virtual organizations. International Journal of Supercomputer Applications 15(3), 200–222 (2001)
6. Wang, S., Zhao, H., Sun, L.: Research on resource integration framework of regional networked manufacturing ASP service platform. China Mechanical Engineering 16(19), 1729–1732 (2005)
7. Cai, Y., Ni, Y., Fan, F.: Research and development on the special ASP platform for the rag trade informationization. Computer Systems Applications 6(2), 14–17 (2006)

# A New Method for Extraction of Signals with Higher Order Temporal Structure

Lei Hu, Bowen Chen, and Zhen Huang

School of Electronic Information, Wuhan University, 430072,  
Wuhan, China  
{ogenius, 691319000, 371234962}@qq.com

**Abstract.** This paper addresses the problem of semi-blindly extracting one single desired signal using a priori information about its higher order temporal structure. Our approach is based on the maximization of the auto-correntropy function for a given time delay. Those values provide information which allows the proposed method to adapt a demixing vector to extract the desired signal without the indeterminacy of the permutation problem in blind source separation. Moreover, this method is different from those for Independent Component Analysis that separate all the available sources, which in some problems, is not desirable or computationally possible. Also, the flexibility brought by the Kernel size selection allows the user to choose the range of statistics he is interested in. We show in simulations that correntropy achieve better or equal separation than other linear methods proposed in the literature for source extraction based on temporal structures.

**Keywords:** Blind source extraction, Correntropy, Information Theoretic Learning (ITL).

## 1 Introduction

The recently introduced similarity function named correntropy[13] can be understood as a generalization of correlation. Correlation measures the similarity between two random variables by comparing their second moment around the mean. When comparing Gaussian variables this is an optimum measure since the higher order statistics are constant for such variables. But this is not sufficient for non-Gaussian signal processing. When comparing any two signals, it would be better to analyze their higher order moments [5], the divergence between their probability density functions [3], or the probability of believing one random variable to be equal to the other. This last method is exactly what correntropy does. Given any two random variables X and Y, correntropy estimates the probability that X-Y equals zero under certain mild conditions [13].

The motivation for our present study of temporal structure using correntropy lies on the problem stated by Barros and Cichoclei in [2]. There, they proposed an algorithm for Blind Source Extraction (BSE) [6] of a specific signal using a priori information about its auto-correlation function. Basically, they calculate:

$$y(t) = \mathbf{W}^T \mathbf{x}(t) \quad (1)$$

$$\begin{aligned}\mathbf{W} &= \arg \min_{\mathbf{W} \in S_1} E[(y(t) - y(t-\tau))^2] \\ &= \arg \max_{\mathbf{W} \in S_1} \text{corr}[y(t), y(t-\tau)]\end{aligned}\quad (2)$$

Where  $E$  is the mathematical expectation,  $S_1$  is the unit sphere and  $t$  is a time delay available from a priori information, for example, by computing the autocorrelation function of the desired source and finding the nonzero delay with the highest autocorrelation.  $\text{corr}$  denotes the correlation between the arguments. In words  $\mathbf{W}$  must maximize the correlation between  $y(t)$  and its previous value  $y(t-\tau)$  for each index. If the desired process,  $y(t)$ , has no temporally structured correlations based on the a priori information, the demixing vector  $\mathbf{W}$  cannot be estimated using only (1). But, if the signal has a higher order temporal organization we propose to estimate  $\mathbf{W}$  by:

$$\mathbf{W} = \arg \max_{\mathbf{W} \in S_1} v(y(t) - y(t-\tau)) \quad (3)$$

## 2 Barros and Cichocki's Method for Blind Source Extraction

The BSE approach proposed by Barros and Cichocki assumes a zero mean and unit variance source vector  $\{\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_n(t)]^T, t \in \mathbf{K}\}$ , where  $\mathbf{K}$  is an index set. In the model, the observed zero mean and unit variance vector  $\{\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_n(t)]^T, t \in \mathbf{K}\}$  results from a linear mixture of the source signals, written as  $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$ , where  $\mathbf{A}$  is a  $n \times n$  unknown matrix. The source signals further satisfy the following relations for a known index delay  $\tau$ :

$$\begin{cases} E[\mathbf{s}(t)\mathbf{s}(t)^T] = \mathbf{I} \\ E[s_i(t)s_i(t-\tau)] \neq 0 \\ E[s_j(t)s_j(t-\tau)] < E[s_i(t)s_i(t-\tau)], \forall j \neq i \\ E[s_l(t)s_j(t-\tau)] = 0, \forall l \neq j \end{cases} \quad (4)$$

For a finite time set  $\mathbf{K}$  with  $N$  elements, the sample estimation of  $\text{corr}$  in (2) is

$$J(\mathbf{w}) = \frac{1}{N} \mathbf{w}^T (\sum_{t=1}^N \mathbf{x}(t)\mathbf{x}(t)^T) \mathbf{w} \quad (5)$$

$$\mathbf{w} = \text{eig}(\mathbf{c}) \quad (6)$$

Where  $\text{eig}(\mathbf{c})$  returns the eigenvector with maximum eigenvalue of its argument. It causes no loss of generality because any nonzero constant multiplying  $\mathbf{w}$  can be canceled by multiplying  $\mathbf{A}$  by the inverse of the constant; this is the scaling uncertainty in Blind Source Separation (BSS) [6]. Maximizing  $J(\mathbf{w})$  with respect to  $\mathbf{w}$  can be proven to be sufficient to extract  $s_i(t)$  in this case.

$$E[y(t)y(t-\tau)] = \mathbf{w}^T \mathbf{A} E[\mathbf{s}(t)\mathbf{s}(t-\tau)^T] \mathbf{A}^T \mathbf{w} \quad (7)$$

$$E[y(t)y(t)^T] = \mathbf{A}E[\mathbf{s}(t)\mathbf{s}(t)^T]\mathbf{A}^T = \mathbf{A}\mathbf{A}^T = \mathbf{I} \quad (8)$$

Unfortunately if  $E[s_i(t)s_i(t-\tau)] = 0$  for some  $i$ , or  $E[s_l(t)s_j(t-\tau)] \neq 0$  for several  $l$  and  $j$ , the matrix  $E[s(t)s(t-\tau)^T]$  will not be diagonal and  $\mathbf{w}$  as estimated by (2) does not necessarily extract  $s_i(t)$ , even if  $s_i(t)$  is dependent on  $s_i(t-\tau)$ . To solve this problem we should exploit the higher order dependencies in the desired signal and we propose to employ correntropy for this purpose.

### 3 Correntropy for Blind Source Extraction

In order to define the correntropy function between two random variables  $X$  and  $Y$ , suppose that the function  $\varepsilon = X - Y$  is a random variable with pdf  $p_\varepsilon(\varepsilon)$ . For a given set of  $N$  independent observations  $\{x(i)y(i)\}_{i=1}^N$ , from  $X$  and  $Y$ , respectively. The correntropy between  $X$  and  $Y$  is defined as:

$$v(X, Y) = E[\kappa_\sigma(X, Y)] \quad (9)$$

$$\hat{v}(X, Y) = \frac{1}{N} \sum_{i=1}^N \kappa_\sigma(x(i), y(i)) \quad (10)$$

$$\kappa_\sigma(x, y) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-y)^2}{2\sigma^2}\right) \quad (11)$$

It can also be proven that (9) is the integration of the joint pdf of  $X$  and  $Y$  along the line  $X=Y$  for kernel sizes  $\sigma$  approaching zero under mild constraints, and that correntropy induces a metric in the sample space that is robust to outliers and has several other properties. We refer to [13] for details. The property that we are interested here is the one that states that correntropy involves all the even moments of the random variable  $\varepsilon$ . This can be verified by expanding the Gaussian kernel function (11) in its Taylor series and substituting it in (9), to give

$$v(X, Y) = \frac{1}{\sqrt{2\pi\sigma}} \sum_{n=1}^{\infty} \frac{(-1)^n}{2^n n!} E\left[\frac{(X-Y)^{2n}}{\sigma^{2n}}\right] \quad (12)$$

$$E[(s_i - s_{i\tau})\kappa_\sigma(s_i, s_{i\tau}) + 2s_j s_{j\tau} \kappa_\sigma(s_i, s_{i\tau}) - (s_i - s_{i\tau})^2 \kappa_\sigma(s_i, s_{i\tau})] > 0 \quad (13)$$

For the specified time delay  $\tau$  that is chosen from  $s_i(t)$ . Note that now we are exploiting generalized correlations for the extraction, as proposed in [12], but, with the difference that now the procedure is optimal in a statistical sense due to direct estimation of error probabilities using correntropy [13].

At last, for a given Mercer kernel  $\kappa_\sigma$ , the estimation of  $s_i(t)$  using (3) is straightforward by the following gradient ascent rule:

$$\tilde{\mathbf{w}}_{k+1} = \tilde{\mathbf{w}}_k + \frac{\mu}{T-\tau} \sum_{t=\tau}^T (\kappa_\sigma(y(t), y(t-\tau))[y(t) - y(t-\tau)][\mathbf{x}(t) - \mathbf{x}(t-\tau)]) \quad (14)$$

The stability and convergence analysis of (13) can be proven to be equal to the methods in [12] using the positive definite properties of the kernel (11). In [9] it was proven that the gradient:

$$\frac{\partial E[G(y)G(y_\tau)]}{\partial \mathbf{w}} = \frac{1}{T-\tau} \sum_{t=\tau}^T G'(y)G(y_\tau)\mathbf{x} + G'(y)G(y_\tau)\mathbf{x}_\tau \quad (15)$$

**Theorem 1.** Assume that the observed data follows the model  $\mathbf{x}(t) = \mathbf{As}(t)$ , where  $\mathbf{x}(t)$  is white and  $\mathbf{A}$  is orthogonal (data preprocessed by PCA) and  $\kappa_\sigma(y, y_\tau)$  is a sufficiently smooth Mercer kernel [1]. Furthermore, assume that  $\{s_j(t), s_j(t-\tau)\}$  and  $\{s_l(t), s_l(t-\tau)\}$ ,  $\forall l \neq j$  are mutually statistically independent. Then, the local maxima of the correntropy with respect to  $\mathbf{w}$ ,  $\Psi(\mathbf{w}) = E[\kappa(y - y_\tau)] = E[\kappa(\mathbf{w}^\top \mathbf{x}(t) - \mathbf{w}^\top \mathbf{x}(t-\tau))]$ , under the constraint  $\mathbf{w} \in S_1$ , include the  $i$ -th row of the inverse of  $\mathbf{A}$ , such that the desired source signal  $s_i$  satisfies

$$\begin{aligned} & E[<\phi''(s_i), \phi(s_{it})>_{H_\kappa} + <\phi(s_i), \phi''(s_{it})>_{H_\kappa}] \\ & + 2s_i s_{it} <\phi'(s_i), \phi'(s_{it})>_{H_\kappa} - s_i <\phi'(s_i), \phi(s_{it})>_{H_\kappa} \\ & - s_{it} <\phi(s_i), \phi'(s_{it})>_{H_\kappa}] < 0, \quad \forall i \neq j \end{aligned} \quad (16)$$

Where by definition

$$\kappa(X, Y) = \phi(X), \phi(Y)_{H_\kappa} \quad (17)$$

(the well-known kernel trick, where  $\kappa$  is a positive definition Mercer kernel such as the Gaussian) and  $\phi$  is a high (possibly infinite) dimensional function in the Hilbert space  $H_\kappa$ ; with inner product  $'$ ,  $H_\kappa$  reproduced by the kernel  $\kappa$ , with first and second order Frechet derivatives  $\phi'$  and  $\phi''$ , which is default to the usual derivative in the finite-dimensional subspace where the data lives.

**Proof.** Assume that the observed data follows the model  $\mathbf{x}(t) = \mathbf{As}(t)$ , where  $\mathbf{x}(t)$  is white and  $\mathbf{A}$  is orthogonal. Making the change of coordinates  $\mathbf{p} = [p_1, \dots, p_i, \dots, p_n]^\top = \mathbf{A}^\top \mathbf{w}$ , we have

$$\Psi(\mathbf{p}) = E[\kappa(\mathbf{p}^\top \mathbf{s}, \mathbf{p}^\top \mathbf{s}_\tau)] = E[\phi(\mathbf{p}^\top \mathbf{s}), \phi(\mathbf{p}^\top \mathbf{s}_\tau)_{H_\kappa}] \quad (18)$$

Evaluating the gradient and Hessian of (18) at  $\mathbf{p} = \mathbf{e}_i$ , assuming the independence assumptions and making a small perturbation  $\boldsymbol{\varepsilon} = [\varepsilon_1, \dots, \varepsilon_i, \dots, \varepsilon_n]^\top$  at  $\mathbf{e}_i$ , we have

$$\begin{aligned} \Psi(\mathbf{e}_i + \boldsymbol{\varepsilon}) &= \Psi(\mathbf{e}_i) + \frac{1}{2} \sum_{j \neq i} \varepsilon_j^2 E[\phi''(s_i), \phi(s_{it})_{H_\kappa} + \phi(s_i), \phi''(s_{it})_{H_\kappa}] \\ &+ 2s_j s_{jt} \phi'(s_i), \phi'(s_{it})_{H_\kappa} - s_i \phi'(s_i), \phi(s_{it})_{H_\kappa} - s_{it} \phi(s_i), \phi'(s_{it})_{H_\kappa} \end{aligned} \quad (19)$$

*Remark 1.* Assume that  $E[\mathbf{s}\mathbf{s}^\top]$  and  $E[\mathbf{s}\mathbf{s}_\tau^\top]$  is diagonal. When  $\sigma \rightarrow \infty$ ,  $E[\kappa_\sigma(s_i - s_{it})]$  (correntropy with Gaussian kernel) equal  $E[s_i s_{it}]$  (correlation) and

(15) equals  $E[s_j s_{j\tau}] - E[s_i s_{i\tau}] < 0$ , which is the condition for reparability of Barros and Cichoclc's algorithm. Our theorem also extends the GABSE algorithm in [16]. In fact we can state the following.

*Remark 2.* Choosing  $\kappa$  properly such that  $G = \phi$  is a pointwise even function, that captures the desired correlation in the data, the gradient in (13) equals (14). The gradient in (14) defines the generalized autocorrelations for blind source extraction (GABSE) algorithm [16]. For a matrix  $x(t)$  of size  $N \times T$ , the computational cost of (14) is  $(3N+2)T$  multiplications,  $T$  calculations of the nonlinearity  $G$  and  $T$  calculations of the nonlinearity  $G'$ . Thus, the computational cost of GABSE is slightly higher than the proposed correntropy method.

*Remark 3.* Choosing  $\kappa$  as the Gaussian Kernel, (15) becomes

$$E[(s_i - s_{i\tau})\kappa_\sigma(s_i, s_{i\tau}) + 2s_j s_{j\tau}\kappa_\sigma(s_i, s_{i\tau}) - (s_i - s_{i\tau})^2\kappa_\sigma(s_i, s_{i\tau})] > 0 \quad (20)$$

## 4 Simulations

We illustrate the validity of correntropy for BSE with three simulations. First with an auto-regressive (AR) series, the second with a sine modulated Gaussian signal, both linearly mixed with noise. The AR series can be separated using only correlations, but the sine modulated noise shows dependencies in the variance that can be exploited only by higher order methods. A third experiment with fetal Electrocardiogram (FECG) illustrates the method with real world signals.

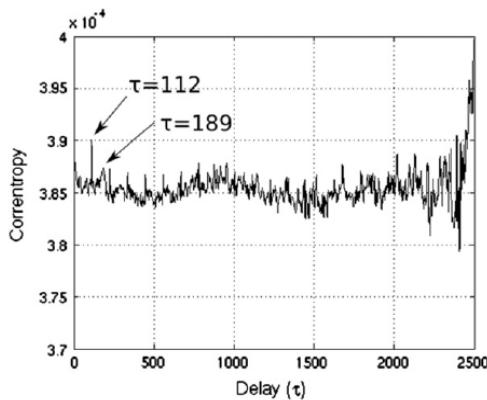
### 4.1 Implementation Details

There are five basic steps for using (13) in practical applications:

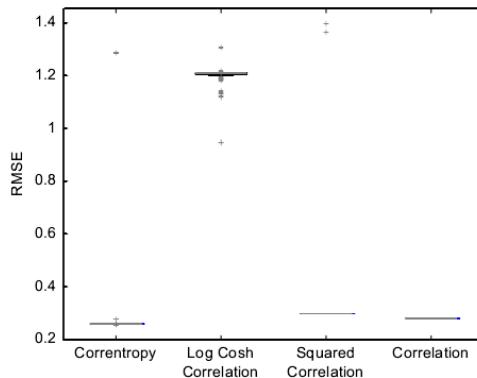
- 1) Pre-process the data  $\mathbf{x}(t)$  using PCA.
- 2) Define a delay  $\tau$  related to the desired signal.
- 3) Define a kernel size  $\sigma$ .
- 4) Adapt the demixing vector  $\mathbf{W}$  using (13).
- 5) Calculate the desired signals by  $y(t) = \mathbf{W}^T \mathbf{x}(t)$ .

### 4.2 Experiment

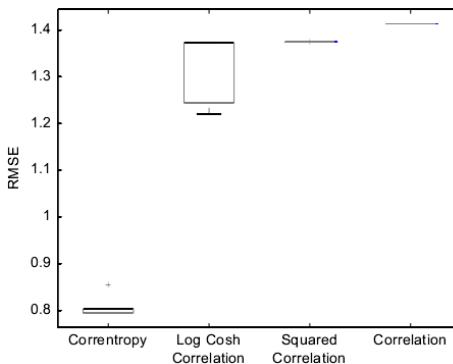
The experiment with FECG is performed with the De Moor's database [7]. In this experiment we must extract the fetal and the maternal component from an eight channel ECG record of a pregnant woman. The delay for the extraction of the fetal component is calculated from the Correntropy function of those eight channels, plotted in Fig1. We see that the correntropy function suggests  $\tau = 112$  and its multiples, as [2] did. We obtained best results for the Correntropy, Log-Cosh Correlation and Correlation methods with  $\tau = 224$  and  $\tau = 112$ , for Squared Correlation and Log-Cosh Correlation method. This illustrates how the delay for extraction varies, in real world problems, with the exploited statistical order of the temporal structure. To extract the desired signal, we used the previous methods and compared their results with the signal separated using



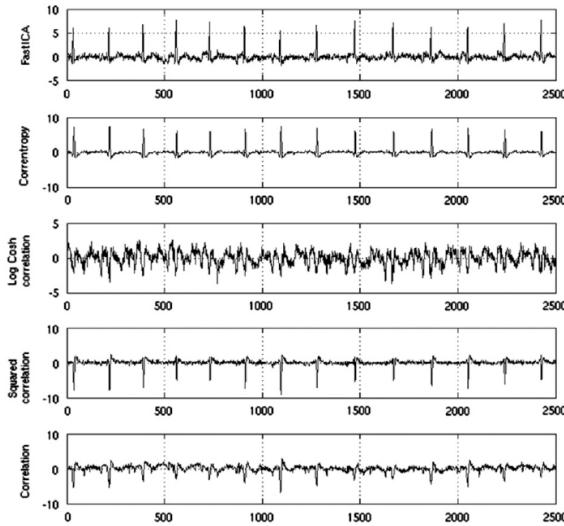
**Fig. 1.** Correntropy function of the first row of the whitened FECG signal. We saw that  $\tau = 772$  is related to the fetal component and  $\tau = 789$  to the maternal component.



**Fig. 2.** Boxplot of the RMSE for the extraction of FECG. For 50 repetitions the gray line is the mean RMSE (Correntropy=0.2591; Log Cosh Correlation=1.209; Squared Correlation=0.2975; Correlation = 0.2799).



**Fig. 3.** Boxplot of the RMSE for the extraction of maternal ECG



**Fig. 4.** ECG signals separated with Fast ICA and extracted by the presented BSE methods

Fast ICA algorithm [9], which will be considered the golden standard, since Fast ICA is a stanlord method in signal separation.

We repeated the extraction of the FECG component 50 times and calculated the RMSE between the extracted component using the compared methods and the signal extracted using Fast ICA. We used  $\sigma = 5$  as kernel size in the correntropy method. The results are boxplotted in Fig2. For  $\tau = 224$  and  $\sigma = 5$  the mean absolute off diagonal values of the correntropy coefficient, correlation, squared correlation and log-cosh correlation matrices are 0.0283, 0.0294, 0.0289 and 0.0312, respectively.

For this experiment the results of correntropy, squared correlation and correlation are similar, with correntropy slightly better than the others. In order to extract the maternal component we used  $\tau = 189$  and  $\sigma = 0.9$  as the kernel size. The RMSE of the extraction using the BSE methods when compared with the Fast ICA extraction are boxplotted in Fig3, and in this case correntropy shows the best performance. For  $\tau = 189$  and  $\sigma = 0.9$  the mean absolute off diagonal values of the correntropy coefficient, correlation, squared correlation and log-cosh correlation matrices are 0.0248, 0.0346, 0.0299 and 0.0371, respectively.

Fig4 depicts the extracted maternal signals. Note that the one extracted with correntropy is less noisy than Fast ICA's (the golden standard), which means that the relative comparisons in Figs 1and 2 should be taken with caution (they assume that the Fast ICA is perfect).

Note that Fast ICA is a blind multiple source separation method which is more computational intensive than BSE [11] and the FECG component must be recognized by the user after extraction since the ordering and amplitude of the sources are unknown after ICA. Alternatively, the correntropy function can be used in BSE with simpler calculations but a priori information to parameterize the method (lag and kernel size) is needed. Thus, the proposed method offers an alternative tool for a unifying approach to source separation [12], since it extracts temporally correlated Gaussian signals and non-Gaussian signals, but we intend to analyze this property in future works.

## 5 Conclusion

This paper proposes the use of correntropy for extracting signals with specific temporal structures. The paper also shows how to recognize the temporal structure of the desired signal using the Correntropy function for extraction. The simulation results showed that correntropy captures generalized correlations for BSE better than other functions proposed in the literature. The simulations showing the validity of the method were based on the extraction of one single desired signal that was linearly mixed with noise. We also tested the method extracting fetal and maternal heart rate in a real database using a single delay. The correntropy method performed equally or better than the ones based on correlation and Independent Component Analysis. On the other hand, the Correntropy method asks for a time delay and a kernel size for correct extraction, which can be cumbersome in some real world problems. Also, if the user does not have a desired signal and no a priori information, it would be better to use a simultaneous separation method such as Fast ICA or a parallel/deflationary implementation of our proposed method [9] and then look any signal that might be "interesting" in the output.

Further studies of the method here proposed should include analyses of the relation between correntropy induced metric (CIM)[1] with the dominant higher order.

## References

1. Aronszajn, N.: Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 337–404 (1950)
2. Barros, A.I., Cichocki, A.: Extraction of specific signal with temporal structure. *Neural Computation*, 1995–2003 (2007)
3. Bell, A.J., Sejnowski, T.J.: An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 1729–1759 (1995)
4. Belouchrani, A., Abed-Meraim, I., Cardoso, J.F., Moulines, E.: A blind source separation technique using second-order statistics. *IEEE Transactions on Signal Processing*, 434–444 (2002)
5. Cardoso, J., Souloumiac, A.: Blind beamforming for non-Gaussian signals, 362–370 (1993)
6. Cichocki, A., Amari, S.I.: *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. John Wiley& Sons Inc., New York (2002)
7. De Moor, D.: Daisy: Database for identification of systems, De Moor, D. (ed.) (1997), Retrieved from, <http://www.esat.kuleuven.ac.be/sista/daisy>
8. Gunduz, A., Principe, J.C.: Correntropy as a novel measure for nonlinearity tests. *Signal Processing*, 74–23 (2009)
9. Hyvarinen, A.: Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 626–634 (1999)
10. Iljin, A., Valpola, H., Oja, E.: Extraction of components with structured variance, neural networks. In: International Joint Conference on IJCNN 2006, pp. 5110–5117 (2006)
11. Leong, W.Y., Liu, W., Mandic, D.: Blind source extraction: standard approaches and extensions to noisy and post-nonlinear mixing. *Neurocomputing*, 2344–2355 (2008)
12. Li, R., Liu, W., Principe, J.C.: A unifying criterion for instantaneous blind source separation based on correntropy. *Signal Processing* 87(8), 1872-1881 (2007)
13. Liu, W., Pokharel, P., Principe, J.C.: Correntropy: properties and applications in non-Gaussian signal processing. *Signal Processing*, 5286–5298 (2007)

# A New Method for Vibration Signal Analysis Using Time-Frequency Data Fusion Technique

Lei Hu, Bowen Chen, and Zhen Huang

School of Electronic Information, Wuhan University, 430072,  
Wuhan, China

{ogenius, 691319000, 371234962}@qq.com

**Abstract.** To overcome the inherent deficiencies of conventional time-frequency analysis (TFA) methods, i.e., different TFA methods or the same TFA method with different control parameters will present different results for the same target signal, a novel scheme named as the time-frequency data fusion (TFDF) is developed in this study by extending the idea of data fusion technique. The TFDF technique can present a more accurate time-frequency presentation for the target signal than what can be achieved by any individual TFA method. Therefore, the TFDF has potential to render a significantly improved time-frequency representation and greatly facilitates extracting time-frequency features of target signals. This will promote the applications of TFA in engineering practices and make TFA methods more acceptable to field engineers. The effectiveness of the TFDF technique is validated by three numerical case studies and the analysis of a rubbing-impact signal collected from a rotor test rig.

**Keywords:** Data fusion, Wavelet transform, Time-frequency analysis, Feature extraction.

## 1 Introduction

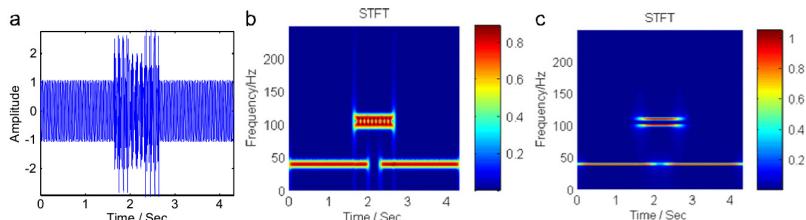
Signal processing techniques [1, 2] have played crucial roles in engineering data analysis. The essential aim is to map a signal from the time domain into another space in which some important information of the signal can be revealed, and consequently, some dominant features of the signal can be extracted. For this purpose, many signal processing methods have been developed, among which the Fourier transform is one of the most widely used and well-established methods partially because of its prowess and partially because of its simplicity. The Fourier spectrum analysis provides a general method for examining the global energy-frequency distribution. As a result, there come some crucial restrictions. The signal must be periodic or stationary; otherwise, the resulting spectrum will make little physical sense [3]. Also because of the global property of the Fourier transform, a shorter periodic event of strong intensity may not be able to give the Fourier spectrum a visible change. Due to the deficiency of the Fourier transform, it is necessary to find more effective methods to analyze non-stationary signals [3, 4].

The radical reason for the difficulties associated with the TFA methods in practical application is that no TFA method can produce the real time-frequency pattern for the target signal (in the real time-frequency pattern, no redundancy component will appear). A TFA analysis result can be regarded as the real time-frequency pattern of a target signal only if the TFA method used to produce the result possesses fine resolution in both the time and the frequency domains. Such an ideal result is referred to as the standard time-frequency representation (STFR) in the present study. If the STFR could be obtained, as this is a unique result, engineers would be able to use the result just like they are now using the Fourier spectrum. Unfortunately, all TFA methods suffer from the restriction of the Heisenberg-Gabor inequality [10], which says that no method can achieve fine resolution in both the time and the frequency domains, and good time resolution definitely implies poor frequency resolution.

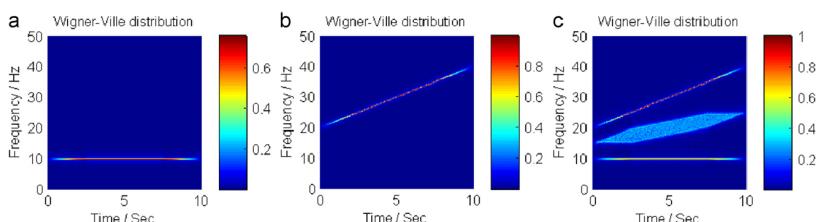
## 2 Time-Frequency Analysis (TFA) Methods

### 2.1 Short Time Fourier Transform (STFT)

To demonstrate the effect of the window size on the STFT, a signal containing three components is analyzed using the STFT method. The three components are: (1) 40 Hz, over the whole time span apart from a narrow disconnection of 0.3 s in the middle, (2) 100 Hz, occupying the middle part with a 1 s duration and (3) 110 Hz, occupying the middle part with a 1 s duration. The sampling frequency is 500 Hz. The temporal waveform of the signal and the STFT analysis results obtained by using different window sizes are shown in Fig. 1.



**Fig. 1.** The demonstration of the effect of the window size on the STFT: (a) the target signal, (b) window size 128 and (c) window size 512



**Fig. 2.** The demonstration of interference terms of WVD: (a) a sinusoidal signal, (b) a chirp signal and (c) the combined signal

## 2.2 Wigner-Ville Distribution

The Wigner-Ville distribution (WVD) [10] is the Fourier transform of the central covariance function of signal. The WVD has excellent concentration in the time-frequency plane. This will mislead the signal analysis. In order to overcome these disadvantages, improved methods have been proposed, such as the Choi-Williams distribution [12] and the cone-shaped distribution [12], etc. However, the elimination of one shortcoming always leads to a loss of other merits, without exception. For example, the reduction of interference term will bring the loss of the time-frequency concentration.

The interference disadvantage of the WVD for multi-component signals can be demonstrated by applying the WVD to analyze two mono-component signals (a sinusoidal signal and a chirp signal) and one multi-component signal generated by combining the two mono-component signals. From Fig. 2(a) and (b) it can be seen that the WVDs have excellent concentrations for the two mono-component signals. However, strong interference term appears for the multi-component signal in Fig. 2(c), and it could mislead the signal analysis as people may regard the interference terms as intrinsic components of the target signal.

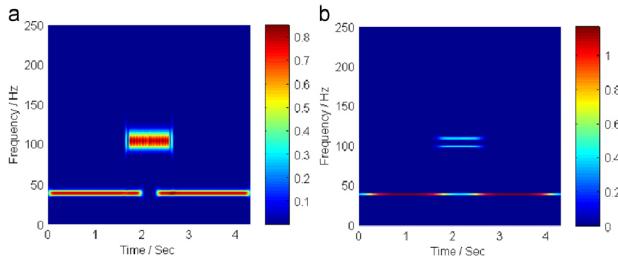
## 2.3 Continuous Wavelet Transform

The wavelet transform (WT) [13, 14] is essentially a kind of window adjustable Fourier transform, which uses a series of oscillating window functions  $\psi_{a,b}(t)$  generated by dilation or compression of a mother wavelet  $\psi(t)$  to scan and translate the signal. The concept of 'adjustable' comes from the following fact. When analyzing the low frequency region, the used window function of the WT is dilated in the time domain, and consequently the WT has a poor time resolution and a good frequency resolution, vice versa. This is the so-called multi-resolution analysis. However, the 'adjustable' doesn't mean 'adaptive', once the wavelet function is given, its resolutions, including the time resolution and the frequency resolution, are determined and cannot be changed. Moreover, the CWT will also suffer from the limit of the Heisenberg-Gabor inequality [10], which states that it is impossible to simultaneously achieve good time and frequency resolutions. In addition, different control parameters could result in different analysis outcomes, even though the same type of mother wavelet function is used. For example, changing parameter  $w_0$  in the Morlet function

$$\psi(t) = \pi^{-1/4} e^{-jw_0 t} e^{-t^2/2}$$

could result in a significant difference to the analysis result.

To demonstrate this, the CWT using the Morlet function as mother wavelet function is applied to the same signal that was previously analyzed by the STFT. The results are shown in Fig. 3. Consequently, in Fig. 3(a), the narrow disconnection of the first component can be detected, while the two components whose frequencies are very close to each other cannot be separated. However, as a larger value of  $w_0$  could improve the frequency resolution while reduce the time resolution, the narrow disconnection cannot be detected in Fig. 3(b), but the two components with close frequencies can be separated.



**Fig. 3.** The effect of the control parameter on the CWT: (a) with small  $w_0$  and (b) with large  $w_0$

### 3 Time-Frequency Data Fusion (TFDF) Technique

The basic idea is to combine data from multiple sensors and related information from associated database to achieve improved accuracies and more specific inferences than could be achieved by the use of a single sensor data alone. Obviously, the idea of data mining technique has a good match with the demand of improving the TFA result through combining the outcomes of different TFA methods. A virtual time-frequency sensor will inherit the properties of an individual TFA method:

First, its input is a one-dimension signal and its outcome is a two-dimensional time-frequency distribution, which is an approximation of the real time-frequency pattern of a target signal. Furthermore, due to the limitation of the Heisenberg-Gabor inequality, either the time or the frequency resolution of the time-frequency distribution cannot be arbitrarily fine.

Second, as all TFA methods satisfy the completeness condition[10] (the basis with which the TFA methods are defined is complete or even over complete), that is, the outcome of the virtual time-frequency sensor always cover the real time-frequency pattern of the target signal. Denoting the non-zero coefficients of the outcome  $TFDi$  of the  $VTSi$  as set  $D_i$  and the real time-frequency pattern of its target as set  $A$ , and the whole time-frequency plane as complete set  $\Theta$ , then it is known that

$$A \subset D_i \subset \Theta \quad (1)$$

Third, the outcomes of different  $VTSi$ s usually are different, i.e

$$D_i \neq D_j, \forall i \neq j \quad (2)$$

It is worth noting here that it is possible  $D_i \in D_j$ , which means that it is possible the time-frequency pattern observed by the  $VTSi$  is contained in what observed by the  $VTSj$ .

Fourth, with the second and third properties, a very useful property is obtained, that is,

$$C = \bigcap_{i=1}^N D_i, A \subset C \subset D_j \subset \Theta, (j = 1, 2, 3, \dots, N) \quad (3)$$

Fifth, as aforementioned, different  $VTS$  are suitable for different types of components in a signal, and if one sensor's attribute matches the attribute of one component very well, then this  $VTS$  can produce an observation of this component with a higher accuracy. Moreover, if one  $VTS$  is a good observation of a signal

component, then the coefficients associated with the component in the TFA result are often considerable.

For all data fusion methods, including the time-frequency data fusion method proposed here, the difficulty comes from the design of a fusion engine. If the target signal is not contaminated by noise, from the fourth and fifth properties of the VTFSS, it is not difficult to design a fusion engine. For example, without loss of generality, assume there are two VTFSSs whose outputs are  $\overline{TFD}_1(t, \Omega)$  and  $\overline{TFD}_2(t, \Omega)$  respectively, then we can easily design a simple but effective fusion engine whose fusion rule is defined as follows

$$\left| \overline{TFD}(t, \Omega) \right| = \begin{cases} \max(\left| \overline{TFD}_1(t, \Omega) \right|, \left| \overline{TFD}_2(t, \Omega) \right|) & \left| \overline{TFD}_1(t, \Omega) * \overline{TFD}_2(t, \Omega) \right| > 0 \\ 0 & \left| \overline{TFD}_1(t, \Omega) * \overline{TFD}_2(t, \Omega) \right| = 0 \end{cases} \quad (4)$$

Or, even more simply,

$$\left| \overline{TFD}(t, \Omega) \right| = \sqrt{\left| \overline{TFD}_1(t, \Omega) * \overline{TFD}_2(t, \Omega) \right|} \quad (5)$$

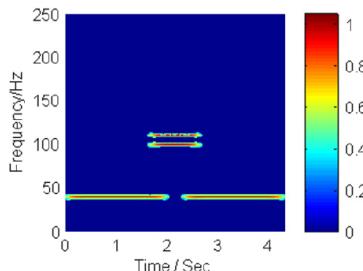
However, in the presence of noise, the situation becomes complicated because the presence of noise would likely make all the coefficients of the outputs of VTFSSs nonzero, which means all sets D are equal, that is,  $D_i = D_j$  for all i and j. Under this condition, the fusion engine of Formula (4) cannot work properly. To remove the effect of noise, a possible amendment to the fusion engine described by (4) is given as follows:

$$\left| \overline{TFD}(t, \Omega) \right| = \begin{cases} \max(\left| \overline{TFD}_1(t, \Omega) \right|, \left| \overline{TFD}_2(t, \Omega) \right|) & \left| \overline{TFD}_1(t, \Omega) * \overline{TFD}_2(t, \Omega) \right| \geq H(\Omega) \\ 0 & \left| \overline{TFD}_1(t, \Omega) * \overline{TFD}_2(t, \Omega) \right| < H(\Omega) \end{cases} \quad (6)$$

## 4 Validation of the TFDF Technique

### 4.1 Numerical Case Studies

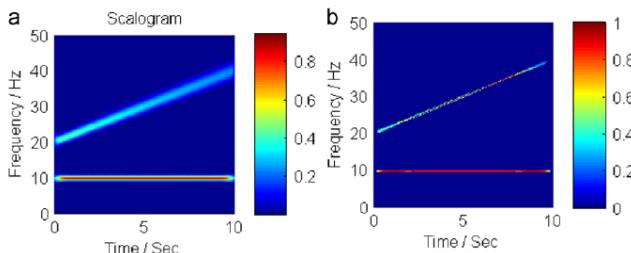
The first numerical experiment is to apply the TFDF technique with the fusion engine described by Formula (4) to combine the time-frequency distributions in Fig. 1(b) and (c) to produce a new TFA result,  $\overline{TFD}(t, \Omega)$ . The result is shown in Fig. 4. Obviously, it has a good resolution in both the time and frequency domain: two high frequency



**Fig. 4.** Using TFDF on STFT results

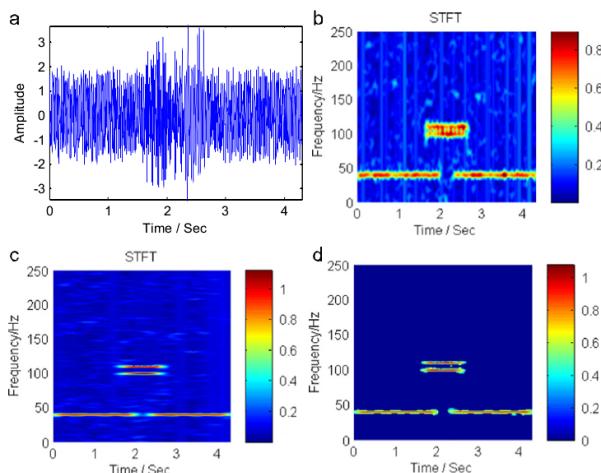
components have been separated so that they can easily be identified and the disconnection of the low frequency component at about 2s can also be clearly observed.

In the second case, a multi-component signal consisting of a sinusoidal signal and a chirp signal is analyzed by applying the TFDF technique to the results obtained by the WVD and the CWT, respectively. The time-frequency distribution determined by the WVD is shown in Fig2 (c) and the result determined by the CWT is shown in Fig5 (a). It can be seen that there are no interference terms in the CWT result, but its concentration is not as good as the WVD result. The new result generated by the TFDF method with the fusion engine defined by Formula (4) is shown in Fig5 (b). Obviously, the new result has a very good concentration and has no interference terms.



**Fig. 5.** Using TFDF on the CWT and WVD results: (a) the result by CWT and (b) the new result by TFDF

In the third numerical case study, the target signal analyzed in the first case is considered again, but in this case the signal is contaminated by a considerable noise. The TFDF technique is applied to two STFT results, which are obtained by using window size=128 and size=512, respectively. Here, the fusion engine (6) is applied, and the threshold is set to be a third of the maximum value of  $TFD_1(t, \Omega) * TFD_2(t, \Omega)$ . The results are shown in Fig6. Clearly, the result generated



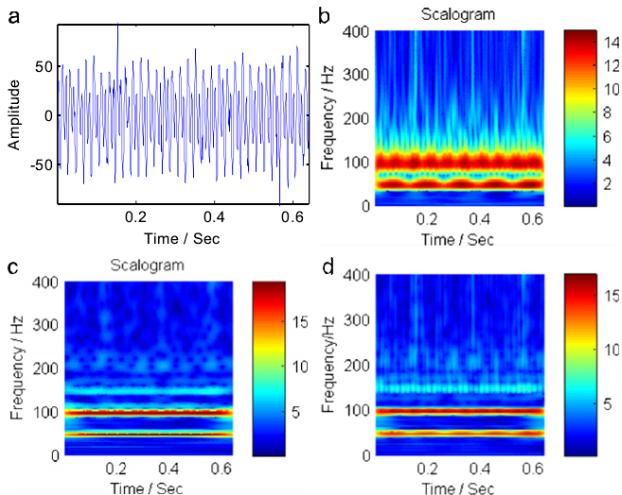
**Fig. 6.** The TFDF for noise-contaminated signal: (a) the signal with noise, (b) the result with window size 128, (c) the result with window size 512 and (d) the result by TFDF

by the TFDF presents a time-frequency distribution of high quality for the noise-contaminated multi-component signal, through which the time-frequency pattern of the target signal can be easily identified.

## 4.2 Impact-Rubbing Signal Analysis

To further demonstrate the effectiveness of the TFDF technique in improving the quality of time-frequency distribution, the TFDF method is applied to analyze one set of vibration signal. The signal was collected from a rotor test rig with impact-rubbing fault by using non-contact eddy-current transducers at a sampling rate of 1.6 1cHz. The rotational speed of the rotor was 3000 rpm.

According to the time-frequency distribution shown in Fig. 7(b), which has a good time resolution but a poor frequency resolution, the amplitudes of the frequency components above 100 Hz change periodically, and the maximal points of these components basically correspond to the occurrence time of the rubbing-caused impacts. However, due to the poor frequency resolution, interference terms appear in the frequency region below 100 Hz in Fig. 7(b). When a large value is used for the control parameter  $w_0$ , the time-frequency distribution calculated by the CWT has a good frequency resolution without interference term, as indicated by Fig. 7(c). Consequently, three main components of frequencies 50, 100 and 150 Hz can clearly be observed in Fig. 7(c). However, the time resolution deteriorates in this case, and so the periodic patterns of rubbing-impacts can no longer be identified through the high frequency components in Fig. 7(c). The result in Fig. 7(d) is generated by the TFDF using Formula (6) as fusion engine. Compared to the results in Fig. 7(b) and (c), the result in Fig. 7(d) presents a much better time-frequency distribution for the rubbing-impact signal where not only the three main components can be clearly identified, but also the periodic patterns of rubbing-impacts can be easily revealed, especially through the frequency component of 150 Hz.



**Fig. 7.** The TFDF for an impact-rubbing signal: (a) a rubbing signal, (b) with small  $w_0$ , (c) with large  $w_0$  and (d) the TFDF result

## 5 Conclusions

In this study, inspired by the idea of data fusion technique, a novel scheme named as the Time-frequency data fusion (TFDF) is proposed to overcome the inherent deficiencies associated with any particular TFA method, i.e., diversity of results, and to render a nearly standard time-frequency representation for target signals. The TFDF works by combining results of several TFA methods to achieve a more accurate time-frequency presentation than could be achieved by the use of an individual TFA method. The effectiveness of the TFDF technique is validated by three numerical simulation signals and a rubbing-impact signal which was collected from a rotor test rig. The significantly improved time-frequency representation achieved by the proposed TFDF technique will promote the use of the TFA methods in practical applications and enable the powerful TFA methods to be applicable in engineering practices and acceptable by field engineers.

## References

1. Jardine, A.K.S., Lin, D., Banjev, D.: A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech. Syst. Signal Process.*, 1483–1510 (2006)
2. Peng, Z.K., Chu, F.L.: Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. *Mech. Syst. Signal Process.*, 199–221 (2004)
3. Huang, N.E., Shen, Z., Long, S.R.: A new view of nonlinear water waves: the Hilbert spectrum. *Annu. Rev. Fluid Mech.*, 417–457 (1999)
4. Wang, C.T., Gao, R.X., Yan, R.Q.: Unified time-scale-frequency analysis for machine defect signature extraction: theoretical framework. *Mech. Syst. Signal Process. Mech. Syst. Signal Process.*, 226–235 (2009)
5. Geng, Z.M., Chen, J., Hull, J.B.: Analysis of engine vibration and design of an applicable diagnosing approach. *Int. J. Mech. Sci.*, 1391–1410 (2003)
6. Teich, M.C., Heneghan, C., Khanna, S.M., et al.: Investigating routers to chaos in the guinea-pig cochlea using the continuous wavelet transform and the short-time Fourier-transform. *Ann. Biomed. Eng.*, 583–607 (1995)
7. Conn, T., Hamilton, J.: Time-frequency analysis of time-varying spectra with application to rotorcraft testing. *IEEE Antennas Propag. Mag.*, 148–153 (2005)
8. Lebaroud, A., Clerc, G.: Accurate diagnosis of induction machine faults using optimal time-frequency representations. *Eng. Appl. Artif. Intell.*, 815–822 (2009)
9. Padovese, L.R.: Hybrid time-frequency methods for non-stationary mechanical signal analysis. *Mech. Syst. Signal Process.*, 1047–1064 (2004)
10. Leon, C.: Time Frequency Analysis. Prentice Hall, Hunter College, New York (1995)
11. Nawab, S.H., Quatieri, T.: Short-time Fourier transform. In: Nawab, S.H., Quatieri, T. (eds.) *Adv. Top. Signal Process.* Prentice-Hall, Englewood Cliffs (1988)
12. Hlawatsch, F., Manickam, T.G., Urbanke, R.L., Jones, W.: Smoothed pseudo-Wigner distribution, Choi–Williams distribution, and cone-kernel representation: ambiguity-domain analysis and experimental comparison. *Signal Process.*, 149–168 (1995)
13. Rao, R.M., Bopardikar, A.S.: Wavelet Transforms—Introduction to Theory and Applications. Addison Wesley Longman, Reading (1998)
14. Peng, Z.K., Chu, F.L., Tse, P.: Detection of the rubbing caused impacts for rotor-stator fault diagnosis using reassigned scalogram. *Mech. Syst. Signal Process.*, 391–409 (2005)

# A Multi-layer Security Model for Internet of Things

Xue Yang<sup>1,2</sup>, Zhihua Li<sup>1</sup>, Zhenmin Geng<sup>1,2</sup>, and Haitao Zhang<sup>1</sup>

<sup>1</sup> Engineering School of Internet of Things, JiangNan University, Engineering Research Center of Internet of Things Technology, Application Ministry of Education, Jiangsu Wuxi, China, 214122

<sup>2</sup> Wuxi Cinsec Information Technology Co., Ltd, Jiangsu Wuxi, China, 214122  
yangxue0223@yahoo.com

**Abstract.** Since the open architecture at the perception layer, the IoT not only faces the traditional security risks in TCP/IP network, but also faces the new emerging risks. In this paper, through analysising the infrastructure of IoT and the security risks in IoT, a new multi-layer security model is proposed, which aims to improve the security technique in IoT. Based on the new model, a series of effective ways to the security risks are discussed at each different layer, respectively. The characteristics of the model and the application domain are summarized to give out, too.

**Keywords:** Internet of Things, Security risk, Security ways, Multi-layer security model.

## 1 Introduction

IoT is another up-and-coming information and technology industry after the computer and Internet. The defination of European Union[1] of IoT is: IoT is a dynamic infrastructure of the global network. It has the self-organizing ability of standard and interoperability communication protocol. The physical and virtual "Things" have identity mark, physical attributes, virtual character and intelligent interface, and are integrated seamlessly with the information network. The understanding of IoT can be understood from technical and application two aspects. We can see that IoT isn't a new technology, but the summary and fusion based on the original Internet. Its basic characteristics can be summarized as comprehensive perception, reliable transmission and intelligent processing. There are opportunities but also challenges[2] during the development of IoT. The biggest problem is the security in the application, that is to say, the development of technology and the moral challenge are both the hot potato. There have been many discussions about the IoT security[3-7], but some discussions just put forward the shortage of IoT mechanism and the others just provide strategic analysis from the framework of macro aspects. We come up with a system architecture to the concrete problem in this paper.

## 2 Infrastructure of IoT

Combining the concept of the IoT, the structure of IoT is divided into three layers: perception layer, network layer and the application layer. They are described briefly as the follow.

**•Perception Layer.** It's the basic and main comprehensive perception part in IoT. It is composed of various types of collecting and controlling modules. Its main function is perceiving and gathering information. For example, temperature sensors, sound sensors, vibration sensors, pressure sensors, RFID and two-dimensional barcode are mainly used to identify the objects. It is similar to the body skin perception part of human bodies.

**•Network Layer.** It is also can be called transport layer. Its work is reliable transmission. It transmits the data through Internet and mobile telecommunication network and so on. Because the perception layer may collect the large amounts of data in real world applications, which making the network layer transport the vast amount of information. It needs certain information processing and management ability.

**•Application Layer.** The main work of this part is to process the data intelligently so that the processed information can be used by us. And then we can get some important real-time information, which is the goal of developing IoT.

## 3 Overview of Security Risks and Security Ways in IoT

### 3.1 Security Risk

Each of the three layers have their own indispensable role and characteristics. The potential security problem can be analysed[8] according to the three-layer structure.

**•Perception Layer.** It is the key basic part of IoT. The research of sensor node is the hot spot both domestic and foreign. At present, since the technical flaws, attackers can easily eavesdrop the communication link, then they can analyze the data and the role of the nodes to capture the users' information. And that may cause great loss.

**•Network Layer.** It is the critical part in the whole IoT system. There may be different structure network connected to each other. The openness characteristic of IoT makes it face many identity authentication security problems. In addition, one of the numerous characteristics of IoT is the huge amount of data. When sensor nodes are perceiving, they produce a large number of redundant data inevitably, which will causes network congestion in the process of transmission. And this is likely to generate denial of service attacks. So we must add the filtration devices between the transmission layer and the application layer to ensure the network unblocked.

**•Application Layer.** As is known to us, application is the purpose of developing IoT. It makes our life become more intelligent and reduces our workload. However, in the application process of a specific industry, perceptive layer collects large amounts of data of users, including some privacy information[9]. Therefore, how to protect these important privacy data involving individual, enterprises or states is the big question of application layer.

### 3.2 Security Ways for IoT

Corresponding to the three important layers of IoT, here, we propose several security ways, based on it, a new security model multi-layer security model for IoT is proposed. The model is shown as figure 1. In which, through discussing the security problem in the process of developing and IoT applications, we propose some security ways. In order to resolve the puzzle of information privacy protection, we add some IoT middleware, such as the third party sever, encryption/decryption mechanism and access control, etc. to establish the improved IoT model on the basis of the original IoT infrastructure.

Based on the above new model, the new security ways for IoT are discussed as the follows.

(1) To strengthen the protection of local domain. Managing and protecting the identity of the tag on some level to make the information of local area can be shared. But the data will be encrypted by 3DES[10] encryption algorithm when they are uplinked to another local area. To ensure the data are encrypted ciphertext when they are transited in each route. 3DES uses three key  $K_1, K_2$  and  $K_3$ . The ciphertext  $y$  of plaintext  $x$  is as

$$y = E_{K_3}[D_{K_2}(E_{K_1}(x))] \quad (1)$$

And decrypting them when they reach to another node and integrating the data with the information of current area. Thus, the information increases gradually between the encryption and decryption with different algorithms. This can effectively prevent the data leaking out in the process of transmission. The process can be described as figure 2.

(2) Making different layers have different permission through authentication constraint of different layer so that we can strengthen the privacy protection dynamics. For example, the electronic medical record card used in intelligent medical and the different data storage database used in intelligent home, etc. These information involves the privacy of users in a certain extent. So we should strengthen data access control authentication, especially the access control connected to the database to avoid active leak or passive leak.

**Definition 1.** Active leak means the information leakage caused by the purposive behavior that some illegal users enter into the related database to obtain the privacy data.

**Definition 2.** Passive leak means the information leakage caused by the accidental behavior that some people just want to access the data which they have the authorization to access, but gain the sensitive data in long odds.

(3) Adding the trusted third party--the third party sever, which can be taken as monitoring module. Before the data access to application layer, they are filtered by the third party to ensure the sensitive data didn't be stolen or interpolated. The sketch map is shown as figure 3.

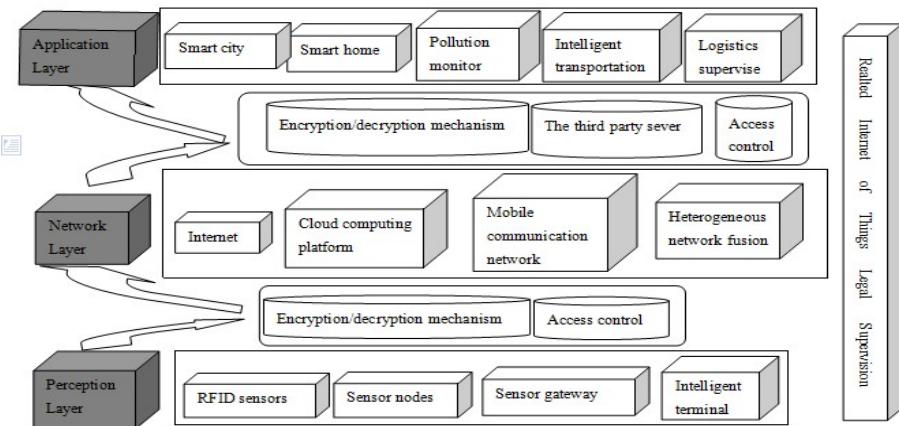


Fig. 1. Multi-layer security infrastructure of IoT

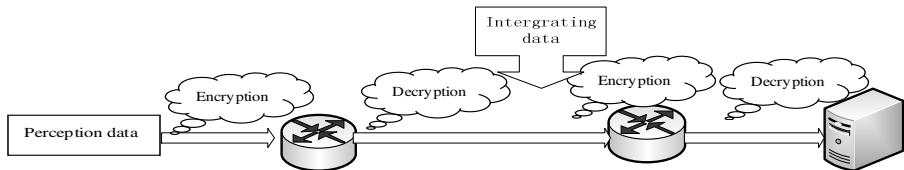


Fig. 2. The schematic plot of encryption transmission

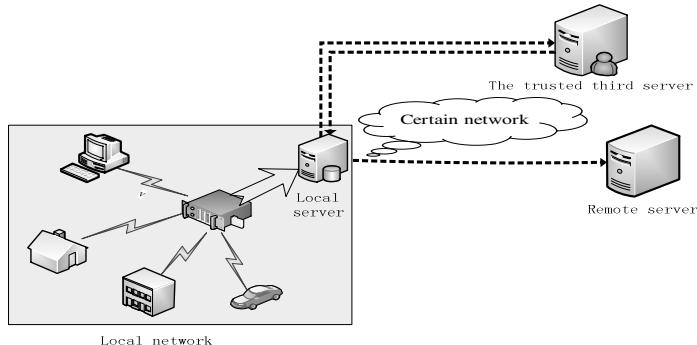


Fig. 3. The trusted third party mechanism

(4) Anonymous processing the data involving the privacy of users according to the security privacy control model. Only those legitimate users passed the authorization can see these processed information. Thus, some users can't see the privacy data after anonymous processing, even though they can access to the system through some illegal path.

(5) The most basic thing is to improve quality of the whole people. To reduce the illegal behavior of getting sensitive data as far as possible. In addition, for the

behavior of using sensors to illegally obtaining information legislation as soon as possible, making it clear to the illegal activities and the cost.

Obviously, the security model can be used in most of the application circumstances, besides the situation that the data is used in real-time. Because this model may reduce the conveying speed to some extent. But this security infrastructure of IoT improves the traditional information transmission mode. From the above figure we can see that it actually increases two layers based on the original IoT model. Its security analysis is mainly from the aspect of the core internet security. But actually IoT is much more complex than the traditional internet. This architecture considers the way of information leakage more comprehensively and specifically during the information transmission.

## 4 Conclusion

The development of IoT has become an inevitable trend in the modern society. On the one hand, this will bring us great opportunity of economic development to promote the society development. On the other hand, the nature of biquitous and intelligent of IoT decides it a relatively open system, which will no doubt emerge information security and privacy leak, etc. Therefore, at the same time we vigorously develop the IoT, how to settle the security work has become the urgent need to address during the development process. Analysis shows that the security structure model proposed in this paper has the feature of security, anonymity, credibility and anti-aggressivity. But due to the technological and policy and some other reasons, these strategies may be implemented after a long period of exploration. And in order to satisfy the need of higher level real-time in many fields, we should balance the security ways and the transmission speed in the future work.

**Acknowledgments.** This work is supported by “the Fundamental Research Funds for the Central Universities((grant No.JUSRP211A41)”.

## References

1. Yang, G., Shen, P., Zheng, C.: The Theory and Technology of Internet of Things. Scinece Press (2010) (in Chinese)
2. Shu, J.: The Discussion of the Security Crisis and the Countermeasures of Internet of Things. Network Security (4) (2010) (in Chinese)
3. Medaglia, C.M., Serbanati, A.: An Overview of Privacy and Security Issues in the Internet of Things. In: Proceedings of the 20th Tyrrhenian Workshop on Digital Communications, Sardinia, Italy, pp. 389–394 (2010)
4. Atzori, L., Iera, A., Morabito, G.: The Internet of Things:A survey. Computer Networks 54(1), 2787–2805 (2010)
5. Weber Rolf, H.: Internet of Things-New Security and Privacy Challenges. Computer Law & Security Review 26(1), 23–30 (2010)
6. Ning, H.-S., Xu, Q.-Y.: Research on Global Internet of Things’ Developments and It’s Construction in China. Acta Electronica 38(11), 2591–2599 (2010) (in Chinese)

7. Liu, Y.-B., Hu, W.-P., Du, J.: Network Information Security Architecture Based on Internet of Things. *ZTE Technology Journal* 17(1), 17–20 (2011) (in Chinese)
8. Gan, G., Lu, Z.: Internet of Things Security Analysis. IEEE (2011)
9. Hui, C.: The Three Shortcomings of Internet of Things: Cost, Security and privacy. *Traffic Construction and Management* (7), 28–29 (2010) (in Chinese)
10. He, D., Peng, D., Tang, X., He, M., Mei, Q.: Modern Cryptography. Post & Telecom Press (2009) (in Chinese)

# Security Research on Cloud-Based Logistics Service Platform<sup>\*</sup>

Fuquan Sun<sup>1</sup>, Chao Liu<sup>1,2</sup>, Xu Cheng<sup>1</sup>, and Dawei Zhang<sup>1</sup>

<sup>1</sup> Information Technology and Business Management Department

Dalian Neusoft Institute of Information, Dalian, Liaoning 116026, China

<sup>2</sup> School of Information Science and Technology, Dalian Maritime University,

Dalian, Liaoning 116026, China

liuchao@neusoft.edu.cn

**Abstract.** In order to solve problems that high cost of establishing logistics service platform, frequent system maintenance and upgrades, and increasingly demanding higher technical of people. We used the idea of infinite expansion of cloud computing and shared infrastructure technology to build logistics service platform based on cloud computing. This paper described the basic framework of logistics service platform based on cloud computing. And proposed a security framework by a research on the security of the framework.. Implementation of the framework can be effective in reducing the cost of establishing a service platform and ensure the data security, which will improve the competitiveness of enterprises.

**Keywords:** cloud computing, logistics service platform, security framework, data security.

## 1 Introduction

With the continuous development of science and technology, logistics companies that use the information technology, has formed the core of information technology to transportation technology, distribution technology, handling technology, storage technology, automation, inventory control technology and packaging technology and other expertise to support the pattern of modern logistics equipment and technology. We have made some achievements in the process of building information technology in logistics, but there are some problems, especially the traditional logistics service platform to build enterprise needs a lot of manpower, material resources, and software between the relatively independent, data cannot be shared, it is difficult to match the rapid growth of diverse requirements of information systems and services.

RFID-based internet of things and cloud-based logistics information technology, which proposed in recent years, can effectively solve the logistics problems. Therefore, in this context, many logistics companies tend to use cloud computing model to build a common logistics platform.

---

\* Fund: Technology Support Program (2009BAH47B06); Higher College Research of Liaoning Province Education Department (L2010046)

Using of cloud computing bring to companies and industries many new opportunities, especially to the logistics companies. but because of reasons of their own cloud computing and network security flaws, resulting in the logistics service platform Security risk. Security problems of Cloud computing platform for applications in logistics services has restricted the development of cloud computing services platform logistics. Therefore, to propose the security framework of cloud computing logistics platform, and to take appropriate security policies to ensure data security is very urgent.

## 2 Design of Cloud Computing Based Logistics Service Platform Framework

In recent years, IBM, Yahoo and Google and other companies are seeking to develop cloud computing technology [1]. Cloud computing is emerging as a shared infrastructure approach, which based on the development of a new computing model, such as distributed computing (Distributed Computing), Grid Computing (Grid Computing), Parallel Computing (Parallel Computing). It will bring a fundamental change in work and business model. In cloud computing, users can use mobile phones, computers and other devices through the network to obtain the necessary hardware, software and other resources, sharing resources, quickly and easily access the services they need. In the user view, resources in the cloud are infinitely expandable, and can be purchased and used at any time.

### 2.1 The Advantage to Build the Logistics Service Platform Based on Cloud Computing Platform in the Enterprise

Cloud computing move the data processing from a personal computer or server to the Internet super-computer cluster with high-speed Internet transmission capabilities, this cluster of computers is very common by the tens of thousands of industry-standard servers, which are managed by the large data processing center.

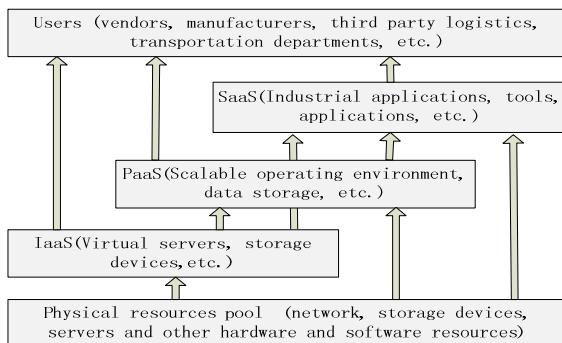
Cloud computing can help companies quickly build a logistics service platform, on-demand purchase and use, thus greatly reducing the cost of construction and maintenance of logistics service platform. Cloud computing needs of lower business equipment, logistics companies can use existing enterprise computer and network equipment; you can enjoy the cloud computing services with very little money. In addition, with the cloud computing service provider's professional help, the logistics companies can maintain the hardware and software systems easily. Such companies can focus more on enterprise's major business.

Cloud computing can help companies quickly and easily at any time for daily business activities. Users can use mobile phones, computers and other devices to take business activities through the network at any time and any place, such as commodity stocks, transportation and other business activities. Employees can even take home to complete the task.

Cloud computing can help companies achieve the sharing of information resources. Various logistics enterprises can take advantage of cloud computing that provides a powerful ability to work together to achieve the sharing of logistics information resources [2]. Cloud computing platform can make dynamic expansion based on the needs of logistics business, sharing the information of many logistics enterprises, reducing the time and money to build a single logistics service platform.

## 2.2 Framework Design of Logistics Service Platform Based on Cloud Computing

Existing logistics business has its own information system, we can't turn them into cloud computing service model one-time, it is an orderly and gradual process. Therefore, under the circumstances, combined with the technologies of cloud computing, SAAS, SOA, etc. A company can use the basic framework of logistics services platform as shown in figure 1.



**Fig. 1.** Basic framework of Logistics Information Platform

### (1) Physical resources pool

Physical resources pool, including networks, storage devices, servers and other hardware resources, these distributed resources connected to the network with virtualization, cloud computing is base of building a logistics service platforms.

### (2) Infrastructure services

By third-party service providers, but also by their own cloud computing center, you can create one or many cloud services center, build their own infrastructure platform IaaS; at the same time, we can appropriately use some IaaS services provided by third parties.

### (3) Platform services

The platform provides the development environment, server platforms, middleware, hardware resources, unified authentication services, billing services, data storage, etc. These service either directly to the user, or can also be provided to the user by a group of Open API. Platform can deploy to the National Center and it can be deployed in some provincial centers to build the logistics service centers based on cloud.

#### (4) SaaS service platform

The platform provides the ultimate application services to the terminal. Such services include the exchange of data between government departments SaaS services, cargo tracking SaaS services, intelligent delivery SaaS services, logistics monitoring SaaS services SaaS services, inventory management, logistics and demand information dissemination and financial management of SaaS services, etc.

### 3 Security Issues of the Logistics Service Platform Based on Cloud Computing

In the traditional logistics service platform, the service providers only provide framework and network architecture, other equipment provided by their own logistics companies, including servers, firewalls, software and storage devices, etc. Companies have full control over physical equipment and software systems, it can be customized the security and reliability of the technology according to their needs. But in cloud computing environment, cloud computing is to provide SOA-based platform for public computing grid computing, the software, computing power, storage capacity as a service to provide public use and collect related royalties. This is a huge change in thinking of the computer world, which will affect the future development of a major innovation, but there are some security risks.

#### 3.1 Security Risks of Cloud Computing Platform in the Logistics Services

Characteristics of cloud computing technology led to the existing security cannot completely solve the security problem, as follows:

- (1) the traditional security domains is invalid; in tradition, through the physical and logical security domain definition we can clearly define the boundaries and protective equipment users, but cannot be achieved in the cloud computing.
- (2) The need for data access control permissions. Users should be able to control who can access their own data. Each of the data needed to visit the user authentication and authorization,,and user access to review the situation.
- (3) Security of user data in storage. The biggest concern of logistics enterprise is that data security of cloud computing. In the cloud computing, most of the business information is stored in the "cloud", the logistics business will not be able to monitor the sensitive corporate information. They worry how cloud computing service providers to ensure that sensitive corporate data is not illegal collection, processing and use.
- (4) Safety of user data in transmission and operation. User data in the run (run-time system data is loaded into memory) and transmission (including cloud computing centers in the internal network and Internet transmission) will not be viewed or changed by others.
- (5) Data migration. When we need to change cloud provider because collapse whether the company's existing data integrity migration to other cloud providers securityly.
- (6) Data integrity. Need to ensure that data remains the same at all times, do not change and damage occurs over time.

In short, due to the complexity of cloud computing, the user's dynamic [3], the cloud computing environment to ensure mutual authentication between the different subjects, each of the main question of trust and confidentiality and integrity of communications, computing availability and confidentiality, so that the cloud computing environment can be applied to the different nature of the safety requirements and stable operation.

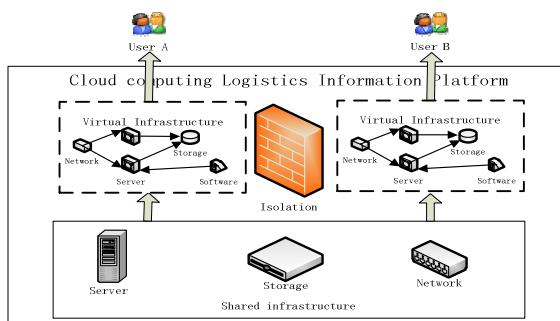
### 3.2 Security Policy Requirements of Cloud Computing Platform for Logistics Service

To solve these problems, the cloud-based logistics service platform should have the following security policy:

- (1) Cloud-based logistics service platform has multiple levels of security domains, each security domain should be the main global and local maps; operation in different security domains to each other identification;
- (2) Communications Security: The physical resource pool and cloud computing platform, between the user and the cloud computing platform through SSL, VPN, PPTP and other security methods to ensure communication security;
- (3) License: There are multiple Authorization in services, owners, agents and users;
- (4) Certification Requirements: Provides a complete single sign-on authentication, agent, co-certification, resource certification, authentication between different security domains complex ways, to meet the dynamic requirements of the user [4];
- (5) Data security: Depending on the user quality of service requirements for data storage confidentiality, integrity, provide different protection, while increasing availability.

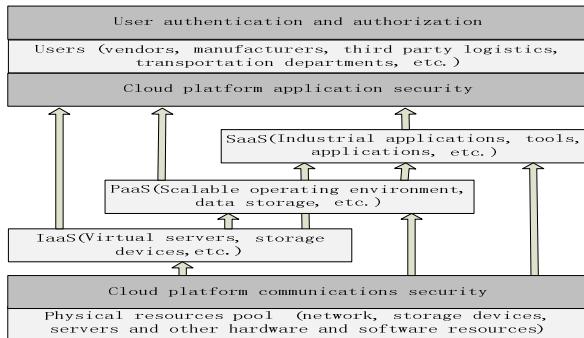
## 4 Cloud-Based Logistics Service Platform Security Framework

Cloud-based logistics platform with unified management of system resources to provide with a variety of resources to services[5] to customers, each customer run relevant procedures under their own virtual resources. Users can control these virtual resources, security policies, and service providers need to ensure that these virtual resources is isolated through certain technical means, as shown in Figure 2.



**Fig. 2.** Schematic diagram of the virtual resource isolation

Basic security services and architecture definition includes logistics enterprise security framework's three core technical infrastructure and related services: user authentication and authorization, cloud platform application security and cloud communications security, from three levels of cloud-based logistics service platform to control security problems. As shown in Figure 3.



**Fig. 3.** Security framework of Logistics Information Platform Based on cloud computing

Based on the information security framework of cloud-based logistics service platform, in accordance with security policy requirements of cloud-based logistics service platform, you can refer to table 1 to realize the framework.

**Table 1.** Implementation table of logistics information platform and cloud security framework

Security framework of Logistics service platform	Cloud services platform security tools
User authentication and authorization	Centralized identity access management
Cloud platform application security	Isolation of data, encryption and protection, parental control
Cloud platform communications security	Network isolation

(1) User authentication and authorization. User authentication and authorization to authorize legitimate users access to systems and data, while protecting these resources against unauthorized access. Through centralization of identity access management, the customers of cloud computing use a standards-based approach to protect the assets that affect productivity and information, and enable enterprises to meet security needs, reduce costs, improve efficiency and avoid risks.

(2) Cloud platform application security. To ensure that cloud-based logistics service platform application security measures include two things: data isolation and classification of security controls and encryption. Customer's data storage of cloud-based logistics service platform can be achieved in two ways: to provide a unified shared storage device, or provide a separate storage device. Shared storage device can store maps and other functions to ensure data isolation; a separate storage device from the physical level of isolation to protect the customer's important data.

Data encryption in the cloud of the specific application form: the client data is encrypted using the user key, and then upload to the cloud computing environment,

and then again when used to decrypt, to avoid the decrypted data stored in any physical medium. Data encryption can use some mature algorithm, such as symmetric encryption, public key encryption [6].

Classification is a way used to regulate the service providers to make user data that through the provider will not be able to obtain an individual, thereby enhancing the safety of operation and maintenance services.

(3) The main cloud platform communication security measures are network isolation [7]. You can use VLAN, VPN, HTTPS / SSL technology to ensure network security and isolation, improving data security.

Of course, in order to ensure data security, cloud-based logistics service platform must have the data backup and restore management functionality. In addition, users of cloud computing services can select multiple cloud computing service providers, and select a different data center location, so that even if a service stops, users can retain their data, continue to run their own business.

## 5 Conclusion

Cloud-based logistics service platform can implement enterprise data sharing framework as the premise, the logistics enterprises need not to purchase a physical device, just follow the needs of rental service, and logistics companies can reduce the business costs in the establishment of funds and staff input. Cloud-based logistics service platform's security framework, respectively, from the user authentication and authorization, cloud platform application security and cloud communicate security platform three levels to control security problems, logistics companies can effectively ensure the security of data, to get the logistics enterprise's security policy requirements of cloud-based logistics service platform. But only from the technical point of view to explore solutions to the cloud-based logistics service platform security is not enough, information security needs of academia, industry and relevant government departments work together to achieve.

## References

1. Chen, Q., Deng, Q.: Cloud computing and its key technologies. Computer Applications 29(9), 2562 (2009)
2. Yu, H.: Cloud-based Logistics Information Platform. Science and Technology Information 1, 443 (2010)
3. Deng, G., Zhang, M., Zhang, Y., Xu, Z.: Cloud computing security research. Journal of Software 22(1), 71 (2011)
4. Guo, Y., Zhang, N., Shang, J.: Cloud computing security framework. Window of the Doctor (7), 62 (2009)
5. Zhu, near: Smart Cloud Computing. Electronic Industry Press, Beijing, p. 227 (2010)
6. Li, H., Li, H.: Key technology and implementation of credible cloud security, 110 p. People Post Press, Beijing (2010)
7. Rittinghouse, J., Ransome, J.: Cloud Computing Implementation, Management and Security. CRC Press, Boca Raton (2009)
8. Heiser, J., Nicolett, M.: Assessing the Security Risks of Cloud Computing (EB/OL) (June 3, 2008), <http://www.gartner.com/DisplayDocument?id=685308>

# Security Technology Analysis of IOT

SheQiang Peng<sup>1</sup> and HongBing Shen<sup>2</sup>

<sup>1</sup> Basic Department Communication Command Academy, Wuhan, China

<sup>2</sup> 6 Series 21 Team Communication Command Academy, Wuhan, China

psqntclass@sohu.com, 573082406@qq.com

**Abstract.** This article reviews the security technology and implementation strategies of IOT from perception layer, network layer and application layer, and solve information security problems of IOT with spread and exploitation process from the technical level, basing on introducing the basic concept, architecture and safety situation.

**Keywords:** IOT, security, strategy.

## 1 Introduction

The concept of IOT was put forward as early as 1999 by Professor Ashton at the American Auto-ID center of MIT. In 2005, ITU gave it the corresponding definition. IOT has become the hot spot question for scientific research technology personnel, because people realized that it has immeasurable potential, such as intelligent power network, intelligent traffic, intelligent logistics, intelligent building, GPS navigation, industrial monitoring, modern agriculture, public security, environmental management, remote medical treatment and digital urban management, digital home, digital battlefield and so on. People expect IOT to bring a lot of convenience in the vision. At the same time, the safety problems such as personal privacy, ethics, laws and regulations, national and military security, hidden in the back of it, should not be ignored. When everything around us are embedded the intelligent chips, everything became so transparent, everyone will think himself living in a state of monitoring, no freedom, no privacy. The expansion application of IOT and controllability of information has become a pair of contradiction. If no the in-depth research, these problems will cause the user to panic.

## 2 IOT and Its Security Technology Overview

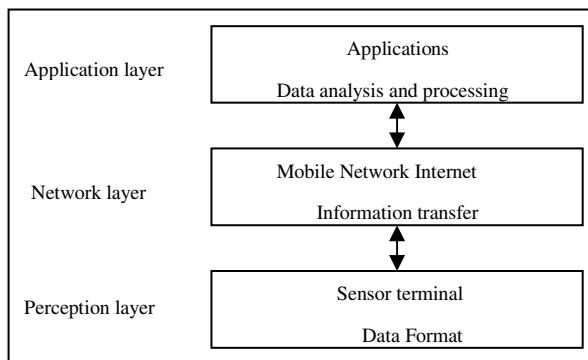
With the developing of the new technology of information, people's desires become more and more intense for the future vision. In these desires, the network technology driven by a thing to things arises at the historic moment, this is IOT. In the early stages of the idea of IOT, people look forward to it can change the production, life style, and give life, work, and study to bring more happiness and joyness.

## 2.1 The Definition of IOT

There are many kinds of expression about the concept of IOT. At present, more accepted definition is: it is a network, through the radio frequency identification (RFID), infrared sensors, global positioning system, laser scanner, information sensing equipment, any object connected to internet can exchange information and communicate, according to the agreement, to realize the intelligent recognition of objects, location, tracking and monitoring and management.

## 2.2 The System Architecture of IOT

IOT combines sensors technology, computer networks technology and intelligent control technology to achieve communication between things. Things networking system architecture is divided into three layers, respectively perception layer, network layer and application layer [1]:



**Fig. 1.** The System Architecture

## 2.3 The Safety Situation of IOT

Any things have no absolutely perfect process until die from arise. When constantly improved, it is inevitable with more drawbacks and IOT is not exceptional also. What we should do is to avoid disadvantages. It brings our life much more convenience and makes our work much more efficient. At the same time, various security problems will show. In the early stages of any technology development, people pay more attention to the technology itself, including its basic theory and practical application. As technology constantly applied, the negative issues with progress such as prominent social problems, safety problems, must cause the attention of people. Then, to a large extent, people put their vision to technical security field. Therefore, some security technology launched. IOT is also like this. At present, IOT has no comprehensive and large area application, and people put more vision on technology itself. Of course, there are many scholars who show intensive attention to the safety of IOT.

From 2008 till now [2], network attacks become greater frequent, computer viruses are endless. According to statistics, new computer virus arise 460000 kinds, and the quantity of infected computer increase 362 million measurements. The proportion of U disk infected virus is 19%, the hang horse website infected virus is 53%, and theft virus accounted for 46%. Trojan threat seriously, 585000 hosts are overseas Trojan control. The zombie network expands continuously and more than 200 hosts are implanted into a zombie program. Websites tampered with increased rapidly to 54000. There are newly increased 4103 vulnerabilities annually, among them ranged attack 3660. "Zero-day attacks" phenomenon to appear, composite virus increase the difficulty of prevention and the zombie nets become the source of online threats from DDos and junk mail. How to construct the use environment of safety is imminent in the complexity of the network.

The research and application of IOT is still in the primary stage, and a lot of theory and key technology remains to be breakthrough, but the core and foundation of IOT is still the Internet, it is the extend and expand network on the basis of Internet, and it is inevitable compatible and inheritance of the current TCP/IP network, wireless mobile network, etc. Therefore, the existing network security system in most of the mechanism can still apply to IOT, and be able to provide certain network security. We can discuss security problems from the development process of IOT, but also need to adjust and add the mechanism of security according to the characteristics of IOT.

## 2.4 The Basic Principle of Safety

According to the system architecture, the safety of IOT can be divided into the perception layer security, network layer security and the application layer security. At the same time, to ensure the safety of IOT, we should be insisted on the following basic principles on the implementation of networking strategy:

- Positive guidance, create safe and healthy network environment.
- Graded protection, key security protection information network and important information system security.
- Active defense, comprehensive prevention and improve information security protection ability.
- Strengthen the management, play the enthusiasm from all walks of life and together build information security system, based on the national conditions.
- Technology research, ensure and promote the informatization development and make sure IOT is controllable.

## 3 The Perception Layer Security Technology

The perception layer is made up of physical equipments; they can access to information from the environment, such as temperature, pressure, and perception of humidity, photo electricity, hull magnetic, and then transfer the information to network. There are some common perception equipment, such as temperature sensors, pressure sensor, humidity sensors, photoelectric sensors, RFID, SIM card, smart CARDS, camera, intelligent chip, bar code, the intelligent machine, GPS, infrared sensors, etc.

### 3.1 The Organization Ability and the Mutual Support Capacity

Energy and computing power of these perception nodes is usually limited, and most of them deployed in the field under unmanned surveillance. They are very easy to destruction and weak of survival ability. Therefore, perception layer should have the organization ability and mutual support capacity. When one or more perception nodes are damaged or failure, other nodes can spontaneously organize and form new perception network to against the destruction of outside, and improve survival ability. At the same time, it can record and report the specific circumstances of the failed node, and generate alarm information to help diagnose and repair them. In the future battlefield, the invulnerability of network to the final victory of war plays a vital role.

According to the characteristics of the wireless sensor network node, in order to effectively manage their each node, we need to design a kind of security routing protocol with anti-destroying ability. There are typical protocols at home and abroad, such as SPINS, TRANS, INSENS [3], etc. Of course, we can also design independently. INSENS of these is a kind of routing protocol facing the wireless sensor network security and tolerating invasion. It can establish safe and effective WSN routing. The whole process has three stages: first, base station broadcast routing request packet; second, each node unicast a routing feedback bag which contains the topology information of adjacent node; finally, the base station broadcast routing table to each node after verified topology information received. INSENS lessons from the SPINS protocol and ensures the integrity and authenticity of the data. Besides using security mechanism of not complex symmetry key password system and one-way hash function, the protocol also transfer complex work from sensor node to the base station, such as the calculation of routing table, and solve the problem of shortage of the node resources.

### 3.2 Security Domain Management

The perception layer is the source of information. When lots around us have been implanted the RFID chip and related products, any equipment with perception capability can access to information of the objects, but the function of the objects is simple and they do not have identification ability and complex security protection ability. Therefore, we cannot ensure the confidentiality of the information to be felt, and the user's privacy is easy to invasion. Perception node information privacy should cause our attention.

How to let the valuable information felt for yourself and not for others? When you meet a stranger who asked you your name, you don't have to answer his questions to protect your personal privacy, although you know. When you come home from work, you can let the cooker automatically cook rice, electric heat preservation through the network, and it should be respond only to your instructions, not others. Therefore, we can refine out a security domain thoughts. Such as the entire people as a security domain, which includes facial features, limbs and organs as sensing nodes, the center of the domain is the human brain, and each node will be able to make a response to external through the Regional Center for authorization. A business unit, a family also can be a security domain, and the sensor of the domain could send perceive information

to network after the agreement of the administrator. One building, a company, a village and so on can be divided into a secure domain, and the administrator of the domain center is responsible for the security of each node. We can set aside a storage space used to define the secure domain in intelligent chip, and it is easy to solve the security problem of perception information, so as to realize that information of the sensor is controllable.

## 4 The Network Layer Security Technology

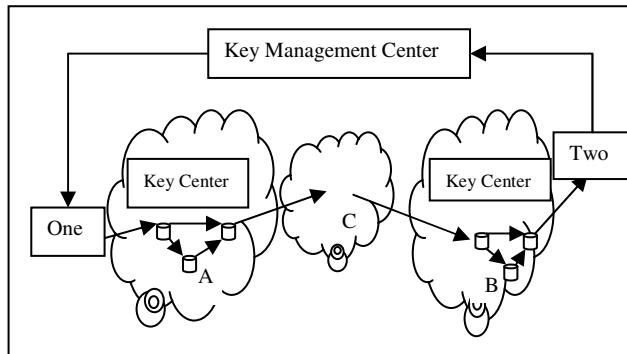
The network layer is mainly used to transmit data, transmits information from perceive layer to destination, and has relatively perfect safe protection ability. With the next generation network technology application, network security protection ability will get the further promotion. But with the development of IOT, this layer will carry the huge data traffic. If no certain security mechanism, the information transmitted on the network is easy to steal, modify, or block, which causes information disclosure, lost, and it is difficult to guarantee the authenticity of the data destination received. Safety problems cannot be neglected. As the main body of data transmission bearing, we should take a simple and effective method to guarantee the security of data transmission. In this layer, as long as we change the strategy, we can solve the secure problem of data transmission with some conventional method.

### 4.1 Security Technology

We can use the traditional security technology to guarantee the security of information transmission of IOT, such as information encryption technology, key management strategy. Password code is the important cornerstone to guarantee the safety of information. Traditional encryption means has two kinds, namely point-to-point encryption and end-to-end encryption. Point-to-point encryption (called semi-encrypt) is an entire packet encryption, including routing information. But the node must decrypt the data received to get the routing information, and then route the data according to address information. Therefore, these nodes which transferred data can get plaintext, and the data in node does not have the secrecy. End-to-end encryption (called all-encrypt) just encrypts data and packets can be routed directly. This method can't protect the source and end of the information being transmitted, and it is vulnerable to malicious attacks from criminals. Although we have the encryption technology, we need to manage key system. Key system is the core of the security, and it is an important means to guarantee the safety of IOT. In order to achieve a unified key management, we can take the way of centralized and distributed. The centralized management of Internet-centric realize the key management of each node of network, through information exchange between key distribution center and each node of IOT, so as to realize the whole key management of IOT; The distributed management of their network-centric is the same way of the centralized management, but each of the hub manage their respective nodes in the network, in order to achieve the whole key management of IOT.

## 4.2 Security Strategy

The traditional encryption technology has its inherent advantages and disadvantages, but we can adjust strategy, optimize the encryption technology.



**Fig. 2.** Encrypted Information Transfer Process

From the figure, we can analyze the process of transmission of information security. This chart uses the distributed key management approach, and uses the combination of the end-to-end and peer-to-peer encryption, the information security, in order to achieve the secure transmission of information. Sender One using Two's public key to semi-encrypt data [routing information + encrypted data], and send it to the A network. A uses the key distributed by its own network key management center to all-encrypt data [cipher text]. When the data is sent over the network from A to C network, A decrypts the data [routing information + encrypted data], and then sent the data to the C network. C processes and routes the data using the methods in its own network. Finally the C network routes the data to the destination network B [cipher text]. B sends the semi-encrypted data to its destination host Two [routing information + encrypted data], B can obtain plaintext using the private key to decrypt the data. From the point of view the whole process of data transmission, ABC networks are not always clear, while the use of the distributed key management approach makes the data being transferred more security. Different networks use different keys to encrypt the data, and the resulting cipher text is still not resolved by routing information and plaintext, even if the data is stolen. It can be seen, such a portfolio strategy is more secure than traditional single encryption policy.

## 5 The Application Layer Security Technology

Intelligent application is the purpose of development of IOT. Through constructing various service platforms, submitting requests to IOT, and getting feedback, IOT shows users a variety of networking applications. These rich applications are also facing security threats, including attacks from network and system, such as system attacks,

identity fake, unauthorized operation, etc. Losses due to these security threats are very heavy every year. In this layer, we can use some existing security mechanisms, such as firewalls, intrusion detection, authentication, access control, etc.

## 5.1 Host System Security

The system security of the host is the first step to guarantee the safety of network application. Hosts on the network all the time may be subject to network attacks. Therefore, we must first solve the security problem of host system, and isolate host system from the external network. Firewall technology is implemented access control method between the public network and internal network. It is a kind of access control measure in the communication between two networks, to allow legal data access to the internal network, and not shut out legitimate data, maximum to prevent network attacks. Within the system of the host, we can use intrusion detection technology to protect themselves. As a proactive security technology, it collects and analysis network, user behavior and system activities, security logs, audits data and checks whether there are behavior violating the security strategy or signs of attack in the network or system, to provide the real-time protection for internal attacks and misuse and intercept harm before compromising network system. Intrusion detection and firewall technology, in combination, can effectively prevent network attacks and enhance the security of the system.

## 5.2 Application System Security

The security of the application system is also very important, because of the foundation of service. In this layer, we can use the authentication mechanism, access control mechanism and other means to effectively prevent unauthorized access, identity fake, unauthorized operation, etc. Through the authentication mechanism, the receiver can verify the real identity of the sender. This authentication mechanism is not just between a men and certification system, and it can be between machines, between programs. To prove your identity, you can use the questions and answers, such as user name, password and authentication code, can also be a mapping relation, such as IP and MAC address binding, still can use the inherent characteristics, such as fingerprints, etc. Now, a popular way of the questions and answers is the user name, password and verification code mode. Verification code is dynamic. Therefore, it is very effective to against the system attacks. To achieve higher authentication mechanism, you can use the combination with the above three ways. Access control mechanism allow legitimate users to access reasonably appropriate resources, prohibit all acts of illegal and unauthorized, and thus effectively prevent unauthorized operation of users, protecting resources and the security of application systems. Here the user can refer to real people and the processes which can access the resources.

### 5.3 Data Security

Information system security, data security is the final analysis, we should ensure that the data is not stolen, destroyed and removed. We can use techniques such as encryption and the like. For example, you can encrypt data for the different user using different encryption strategy and different password, in the operating system or database level. Of course, it will influence the work efficiency of the system, if we encrypt large amounts of data, and reduce the performance of the system, to some extent. But in the future of clouds computing to support desktop system, the technology will provide a broad application space.

## 6 Conclusion

The security technology foundation of IOT is in the protocol itself. If we take fully the security technology widely used at present into account in related protocol of IOT, and provide some alternative configurations, it will bring us more convenient to promote the use of IOT for future. This paper starts from the system architecture of IOT, and put forward some ideas for the use of related techniques, but some questions still need in-depth research.

## References

1. Han, T.: Architecture of IOT. Zhejiang University. Institute of IOT (January 3, 2011),  
[http://www.wsnccs.zjut.edu.cn/article\\_show.asp?Id=256](http://www.wsnccs.zjut.edu.cn/article_show.asp?Id=256)
2. Chen, M., Wang, S.: The Emergence and Development of IOT. Beijing Normal University, 1672-5913, 12-0001-03 (2010)
3. Deng, J., Rham, Mlshra, S.: INTRSN: Intrusion-tolerent rounting in wireless sensor networks. In: Proceedings if the 23rd IEEE International Conference on Distributed Computing Systems (ICDCS 2003), Providence, RI, pp. 65–71 (2003)

# A Study on Mobile Phone Security Industrial Ecology

Qiyu Chen<sup>1,2</sup>, Jinlong Hu<sup>1</sup>, and Ling Zhang<sup>1</sup>

<sup>1</sup> Communication and Computer Network Laboratory of Guangdong Province,  
South China University of Technology, Guangzhou, China

<sup>2</sup> Guangdong General Research Institute of Industrial Technology, Guangzhou, China  
chenqy@scut.edu.cn

**Abstract.** With the rapid increase of mobile phone applications, its security becomes a hit issue recently. The chain of mobile phones is composed by numerous manufacturers, services providers, consumers, and the government. They are also the participants of the mobile phone security industry. An ecosystem of mobile phone security industry has been established. Lots of research was focused on the single security problem of mobile, while the relationship among them was ignored. This paper focuses on some key characteristics of the mobile phone security and analyses the ecosystem environment which the industry faces. It also discusses how to keep the ecological balance of the mobile phone security industry.

**Keywords:** mobile phone security, industrial ecology, ecosystem.

## 1 Introduction

With the wide application of smart mobile phones, mobile phones become both the information center and computing center of people all over the world in recent years. From the statistic data of ITU (International Telecom Union) announced in January 2011, more than 2.08 billion people surf the Internet and the population of mobile phone users was much higher, amounting to 5.28 billion all over the world. At the same time, all kinds of security problems of using mobile phones appear, such as privacy violation, spam, radiation, information disclosure and etc.. Both the government and mobile security vendors agreed that the mobile security has become a significant issue faced by society. These problems look mass; they happen at different nodes of the chains of mobile phones industry and should be analyzed comprehensively.

In these years, ecosystem analysis method has been regarded as an effective tool to solve the complex problems, in the field of both society science and natural science. In the most general sense, an ecosystem of services is a complex system of services, consisting of services and the relationships among them. This perceived assemblage of interrelated services comprises a unified whole to facilitate the flow of resources such as knowledge, competences and added-value [1]. This paper would analyze the security problem of mobile phone and the relevant industry with some key rules of ecosystem. We hope to draw an overall outline of the mobile phone security industry.

## 2 Two Viewpoints on the Mobile Phone Security

### 2.1 From a User's Perspective

Mobile phone security is not a new problem. It appeared when 1G mobile phone was used in 1970s. People tend to analyze the mobile phone security from a user's perspective. There are four major threats. The first is malicious software, such as Trojans, eavesdropping software. Frost & Sullivan (a famous market research institute) predicted in the early 2010 that the number of mobile phones viruses and malicious software will reach up to 2200 or so in China by the end of 2010. The fact is that it is still growing. Malicious software can bring users various hazards. Some will cause economic losses to the users by making an order, sending text messages, doing malicious payments without the user's agreement.

The second category is the malicious harassment, such as spam messages, harassing phone calls, spam, and so on. For 3G networks provide high bandwidth to the malicious harassment, it is more convenient for mobile advertisements to pop up, such as the phone's desktop pop-up ads, favorite bookmarks, and adding a specific phone number.

The third category is the loss of privacy. Because more and more personal information is stored on the phone, the loss or disclosure of personal information can cause immeasurable damage to users.

The fourth category is the radiation of mobile phone. A cell phone radiation and cancer study on normal mice indicated that the animals were partially restrained although these experiments did not reveal mobile phones' promotional or co-carcinogenic effect on skin tumorigenesis as some studies suggested [2]. In addition to restrictions on scientific research, the interests of disputes among the communications industry, government, research institutions, is also an important reason why the conclusions isn't made yet.

While the above categories tell us the general security problems users are facing, it is not the whole story. There are other security problems in the different nodes of the ecosystem chain, such as production safety, battery pollution and etc.. On the other side, it is very difficult for people know the relationship among the security problems. Hence people are always at a loss when the threat appears. In addition, the unclear relationship does harm to the development of the industry, because it can not give the investors clear concept of the industry. Many investment opportunities are hidden, while the current security services can not meet the security needs.

### 2.2 From Industrial Ecology Perspective

Natural ecosystem is an overall system composed of species, which live in the same area with interrelated, biological communities and the local inorganic environment. There are four major components within an ecosystem. They are non-biological environment, producers, consumers and decomposers.

The importance and urgency of the mobile phone security is growing and the market is still at the early stage, but it has great potential future. Who grasps the opportunities earlier can take advantage of the nice situation in future competitions. With the cultivation and rapid growth of the mobile phone security market, the

single-handed security service can not meet the needs in the future. It requires all of the participants work together to keep the balance of the ecosystem.

As an ecosystem, the mobile communications industrial ecology is also composed of four essential components, including social environment, service providers, users and recyclers. Each participant has to face its own security problems, which is shown in the chart below.

**Table 1.** The Mobile Phone Security and the Mobile Communications Industrial Ecology

Components of ecosystem	Mobile industry	communications	The mobile phone security
non-biological environment	social environment(government, policy)	contents security	economy
producers	service providers(mobile phone manufacturers, end sets vendors, telecom operators)	material security, security, service guarantee, code security	production
consumer	users(people, organization)	mobile phone radiation, privacy, malicious software,	malicious harassment.
decomposer	recyclers	waste pollution	

### 3 The Mobile Phone Security Industrial Ecology and Its Characteristics

In the past, the companies make their own efforts in their respective field, being only concerned with their own business. For example, anti-virus software vendors only focus on the application of anti-virus software, while telecom operators are only care about the security of the system platform. However, this model of operation can not meet the security needs of future mobile phone security industry. As the components of the ecosystem, telecom operators, anti-virus software vendors and handset manufacturers should work together, to form the industry chain, to make the ecosystem work. Within the ecosystem, primary rules should be followed. Ecological balance, interdependent species and mutually constraining, material recycling, interaction between biology and environment are the basic rules within the ecosystem of mobile phone security industry too.

#### 3.1 Ecological Balance

Ecological balance of the ecosystem mainly refers to the relative stability among the various species. The stability of the ecosystem depends on the species' survival and development, and their abilities of building the relationship among species properly. A specific environment can accommodate certain number of creatures because both space and resources are limited. When the capacity is close to the saturation, the growth rate will decline if the population increases, even to negative rate, to reduce the number of populations.

**Table 2.** Analysis of Mobile Phones Security Industry

rules of ecosystem	key characters	key words of mobile phone security industry	
ecological balance	Less significant changes in steady environment. Population remains stable.	industrial intelligent protection.	standards, property
interdependent species and mutually constraining	Any species keep relationship and mutual restraint with others within a biological community. Any species have their own niches.	personalized differentiated cooperation.	service, services,
material recycling	Metabolism provides the material necessary for life constantly recycling in an ecosystem.	material recycling, sustainable service.	
interaction between biology and environment	Biological evolution is a product of the interactions between biology and environment.	Steady society, economic and political environment, international relationship.	

Because the domestic mobile phone security market is limited, it will reach saturation a few years later. So some participants are aiming at the oversea market, to keep the ecological balance of domestic market. To do that, the participants should pay more attention to the international industrial standard, so they can provide available security service in foreign countries. It is to say, to keep the ecological balance of industry, one of the effective ways is to enlarge the market. On the other hand, good relationship among species can lead to ecological balance to a better extent. As to the mobile phone security industry, nice intelligent property protecting environment is helpful to provide more opportunities to the participants.

### 3.2 Interdependent Species and Mutually Constraining

Species tends to keep harmonious relationship and mutual restraint within a biological community. Food chain, competition and mutualism are the mainly discussed.

In the food chain, the proportion of the number of species living in two adjacent areas remains relatively stable. Competition among species often happens because species use the same resources. As we know, plant competition for space, water and soil nutrients. Animals fight for food and habitats, and so on. In the long-term evolution, the competition will promote the differentiation of ecological characteristics of species, to ease the competition among them, and to build a certain biological community structure.

In an ecosystem, the species whom are needed most can reach stabilization easier than others. In order to live better in the ecosystem, one should make great effort to retain its advantageous factors, and to overcome its shortcuts in the co-evolutionary process.

One of the most obvious characters of network economy is personalization. With the advent of new technology such as Web 2.0, which enables customers' integration into product design and configuration inexpensively and rapidly, Kotler's observation is beginning to prove a reality—morph customization into personalization of products and services [3]. Facing the different demands of thousand of customers, the

companies have to do their best to meet individual needs. In the competition of mobile phone security industry, the companies should emphasize the differentiated services to find the opportunity.

Finding the appropriate niche in the ecosystem is another important strategy. In an ecosystem, each species has the most suitable location for its own survival time (called niche) in the long-term. Niche is the most suitable position in an industry eco-environment for both a business enterprise and the industry. In the ecological system, for each "species" or "individual", the best way to maintain a competitive advantage in the system is to find the right niche [4]. Besides, mutually constraining is another effective way to keep balance of ecosystem. Participants like to strengthen the cooperation across the chain sectors and create a composition of forces. The sectors on the chain should enhance cooperation, work together, share resources and complement each other, and jointly promote a healthy and rapid development of mobile phone security industry.

### 3.3 Material Recycling

Metabolism provides material necessary for life and is constantly recycling in an ecosystem. Generally speaking, metabolism obeys the rule of "more in and less out" in the beginning, and it reaches a balance when the community tends to mature. People should remember this law. Firstly, people ought to do rational exploitation of biological resources, rather than exploit simply for the moment. On the other hand, people should control the environmental pollution, as a large quantity of toxic industrial waste has been discharged into the environment, beyond the self-purification capacity of the biosphere. As a result, the toxic accumulation does harm to the living environment of human and other organisms.

The number of old mobile phones which were thrown away annually is more than 100 million. Old mobile phones have become a new electronic pollution source. A used cell phone contains at least 20 substances which are potentially pollution. The pollution of a used cell phone battery is equivalent to 100 ordinary batteries, which can contaminate 60,000 liters of water—equaling the volume of three standard swimming pools. In China, the situation of mobile phone recycling is serious, while the ecosystem of mobile phone industry is not stark. We need to accelerate the forming of cell phone recycling system in order to compensate for the lack of industry. In mobile phone security industry, material recycling should be regarded as a very important business from now on. Sustainable service is another idea to extend the life of mobile phones. The users can use the phone longer with the increasingly quality of mobile phones. Provide sustainable service to users can make the users delay changing their mobile phones.

### 3.4 Interactions between Biology and Environment

Biological evolution is a product of the interactions between biology and environment. With the expansion of the field of human activity, the human's behaviors impact on the environment more obviously.

In the mobile phone security industry ecosystem, non-biological factors mainly refer to the industry-related political, economic social and cultural environment. The

establishment of these non-biological factors significantly depends on the government's efforts to develop and implement legislations and regulations. Achievement from the steady social environment, better industrial law and sustainable international relationship will promote the mobile phone security industry.

Based on the analysis above, we can propose a framework for the ecosystem of mobile phone security industry. The ecosystem is composed mainly with manufacturers, service providers, users and recyclers. The economic, social, cultural and legal environment is its non-biological environment. At the different phrase of the lifecycle of mobile phones, there are security problems respectively. Each security has its own niche.

## 4 Conclusion

With the rapid increase of the users of mobile phones, the market of smart mobile phones has been expanding all over the world. Facing the complex phenomenon of the mobile phone security industry, ecosystem is an effective way to help us to clarify their relationship. Mobile phone security industry has a wide market prospect with fierce competitions. The participants should focus on the opportunities according to both the niche of security service and their own advantaged field. The mobile security vendors should aim to provide integrated services to mobile users through continuous product innovation.

## References

1. Biennier, F., Aubry, R., Badr, Y.: A Multi-dimensional Service Chain Ecosystem Model. In: Vallespir, B., Alix, T. (eds.) APMS 2009. IFIP AICT, vol. 338, pp. 563–570. Springer, Heidelberg (2010)
2. Lin, J.C.: Cell-Phone Radiation and Cancer Studies in Normal Mice. IEEE Antennas and Propagation Magazine 51(2), 186–188 (2009)
3. Kumar, A.: From mass customization to mass personalization: a strategic transformation. Int. J. Flex. Manuf. Syst. 19, 533–547 (2007)
4. Huang, Y., Cao, L.-J.: Telecommunication Industry Ecosystem and its Development Measures. Future and Development 3, 32–35 (2011)

# **Research on Sensor-Gateway-Terminal Security Mechanism of Smart Home Based on IOT**

Fei Li, Zhou Wan,  
Xin Xiong, and Jiajun Tan

Kunming University of Science and Technology,  
650500, Kunming, China  
1f8709@126.com, ynkgwz@yahoo.com.cn

**Abstract.** Along with the rapid development of the smart home based on the internet of things (IOT), security issues become more important. On the basis of study of the smart home based on IOT at home and abroad, this paper discusses some security issues and corresponding solutions about Sensor network, home gateway and the application terminal, and research on Sensor-Gateway-Terminal security mechanism. Through organic integration of security about the three parts and making them coordination, we can achieve the safe protection of information transmission and user privacy, and then do the protection of the smart home security on the greatest degree.

**Keywords:** IOT, Smart Home, Sensor-Gateway-Terminal, Security Mechanism.

## **1 Introduction**

The smart home based on IOT is a house as a platform, consists of the construction equipment, information appliances, equipment automation and network communication, which has environmentally friendly and comfortable, safe and efficient living environment with the functions of system, service, structure, management[1]. The smart home system can make all kinds of various subsystems about home life combined and connect the Internet to monitor and manage the exchange of information and communication through sensing devices, which achieve the intelligent home ultimately.

When the smart home based on IOT grows at a high speed, the hidden security issues also gradually expose [4].To ensure the security of the smart home is necessary to meet the physical security, security information collection, integrated security of information transmission and information processing security. The ultimate safe goal is to ensure information confidentiality, integrity, authenticity.

According to the characteristics of the smart home based on IOT, through studying on the internal sensor network, the home gateway and the application terminal this paper raises Sensor-Gateway-Terminal security mechanism. The Sensor is the internal

sensor network. The Gateway is the home gateway. The Terminal is the application terminal. The following paper describes in detail about Sensor-Gateway-Terminal security mechanism.

## 2 Sensor Network

### 2.1 The Analysis of Security Issue

Wireless sensor networks is the internal network of the smart home system, whose task is fully aware of the home environment, collects and passes information to the external network.

In the home wireless sensor network, the prominent security issues is mainly the security issues of sensor nodes, because the information must be collected and transmitted by the nodes [2]. Sensor nodes are miniature devices, have a simple function, which limit their computing power, storage capacity and communication ability. So it is hard to design a complex security agreement and against DOS attack. If a sensor node produces inaccurate data or false data because of dysfunction, then it is easy to lose or confuse the packet data through its transmission. And the intruder may increase own nodes, enter false data and plug the message transmission in the sensor network. The false node which has a strong computing power can pretend to be a sensor node. If a captured sensor node compromises (communication between nodes within the sensor network and the shared key of telemetric platform), the intruder can control the sensor nodes, and obtain the transmission message through the nodes.

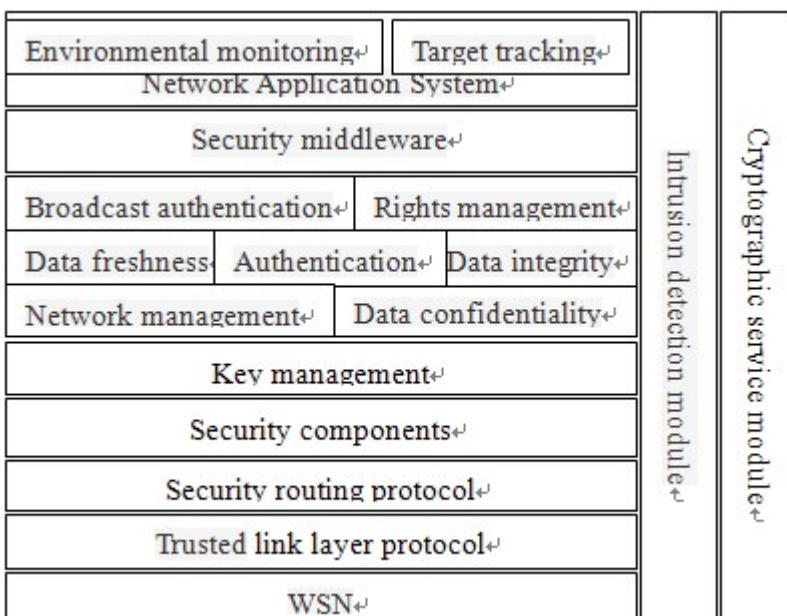
Packet transmission route of the sensor network is a connectionless routing. The channel error, unsecured wireless communication channel, the collision and the time delay may result in the loss of packet transmission.

Also the attacker may undermine the tasks undertaken by the sensor network through causal analysis, information communication mode or the information exposed by sensor. And the unencrypted addressing information and routing information is easy to be attacked by traffic analysis.

### 2.2 Safety Measures

For wireless sensor networks the safety problem of the sensor network need to be considered. It can't be avoided that the security issues of sensor nodes are easy to physical manipulation, so this need to improve other technologies to improve the safety performance of the sensor network. Using key management mechanism is effective within the sensor network. When in communication establishing a temporary session key can improve confidentiality, and the authentication can be solved by non-symmetric cryptography or symmetric cryptography program. The authentication

scheme using non-symmetric cryptography usually requires strong computing power and communications capabilities of the sensor network, and has higher security requirements; the authentication scheme using symmetric cryptography usually need to preset the shared key between the sensor nodes, and it is very efficient and uses fewer resources of sensor nodes, so general sensor networks use this program. It is necessary to complete key agreement and establish a session key on the basis of certification. The solution like secure routing, connectivity can be relatively independent. Secure routing protocols, and intrusion detection is necessary to improve the safety performance of the sensor network. Security framework is shown in figure 1.



**Fig. 1.** Security framework

### 3 Home Gateway

#### 3.1 The Analysis of Security Issue

The home gateway can be said to be an integral part of the smart home system, the main task is to achieve information sharing and conversion between different communication protocols of the home network, and make data and information exchange with external networks. In other words, the family home can achieve to

network between the home smart devices, as well as interconnect with external communications network through the home gateway. The security of the smart home system not only contains to the security of traditional network, but also the security of the appliances. If there are safety risks .the home will have huge losses. So the security is an urgent need to be protected in the smart home system. The smart home system has the following security issues [3].

Posing: an attacker could pose as the end hosts for issuing control commands to the home gateway, or pose as the home gateway for passing false information to the terminal host.

Replay: the replay attack can be divided into replay of the terminal host and the replay of home gateway.

Data theft: getting the transmitted information and data between the terminal host and home gateway through intercepting the data packets, tapping the line.

Virus attack: the attacker adds the virus to the data packet, then releases in the system, takes up system resources wantonly through constant self-replication, so that the system can't complete the relevant work, and it can make the system unusable finally.

Denial of service attacks: the will organize vast amounts of data to access the home gateway at the same time. In addition to the verification of user legitimacy, the server can't complete normal data access, and then can't work.

Illegal processing of user data: the attacker modifies the stolen data, and sent the error message to the home gateway or the end hosts.

### 3.2 Safety Measures

It is vital to the safety of the home gateway, this paper discusses security measures based the layered network [3].

The link layer can be divided into virtual LANs, link encryption communication and other technology programs, which can ensure the confidentiality and integrity of the transmitted data.

The security of the network layer needs to ensure that the system provides the authorized service unless the user has permissions. So the network can avoid the destructions like eavesdropping the information, tampering, posing, and adding illegal messages. We can achieve the security of the network layer through the network security services and access control, such as limiting the IP address, encrypting network layer and using the firewall.

About the security of transport layer, UDP transport protocol uses connectionless secrecy transmission of data. It is vulnerable to replay attacks because it does not need to establish a connection when communicating. The solution is that using identification

mechanism in the other layers to prevent replay attacks at the same time. TCP protocol can effectively prevent simple replay attacks, but the attacker can replay the entire connection process to complete the attack. The solutions of completing the safe transmission of data need to rely on the protocol of other layers, or increase the SSL protocol.

The security measures of application layer can be used for the safe program of the application layer protocol and other security mechanisms such as data encryption, authentication and digital signatures can be implemented simultaneously in order to improve the security of application layer.

Security framework is shown in figure 2.

Security policy (Mechanism of dynamic access)		Mechanism of security policy	
Access control		Mechanism of access control	
Authentication			
Access strategy review			
Equipment reliability inspection			
Secure communication channel	Virtual private network	Mechanism of network security	
Appliances	Remote access users		

Fig. 2. Security framework

## 4 Application Terminal

### 4.1 The Analysis of Security Issue

The terminal equipment can be broadly divided into two types, one is the traditional computer client, and another is a mobile terminal. They are intelligent processing platform of information processing, have functions such as monitoring, controlling, and often carry a lot of important information [5]. Therefore, the safety issues of terminal equipment can't be ignored.

System information processing platform is mainly reflected in the smart, and can process the massive amounts of data efficiently and automatically. The automatic process of judging the malicious data is limited. Intelligent can filter and judge through certain rules, so the attacker can easily avoid the rules and achieve the attack effect. When intelligence makes mistakes, how to reduce the losses caused by attacks to a minimum, and how to return the normal working condition as soon as possible is a major security challenge of the smart home system. Smart systems need to set different access rights according to different application requirements, and different permissions accessing the same database can be obtained different results.

In addition, the temporary intermediate variables generated in the course of implementation of the cryptographic algorithm or cryptographic protocols is not useful after the end of the process, but if the attacker gets these intermediate variables which provide important parameters, so that the attacker can achieve success. So these intermediate variables need to be destroyed timely.

With the network of universal access to information, more and more people attention to user privacy. The user information is required information on the certification process, so the information privacy is a hard challenge.

## 4.2 Safety Measures

Along with the rapidly growing popularity of the smart phone, it is used more and more as the application terminal. In view of the above security issues, this paper makes the following precautions about the smart phone.

Phone starts in Safe mode with the following two steps.

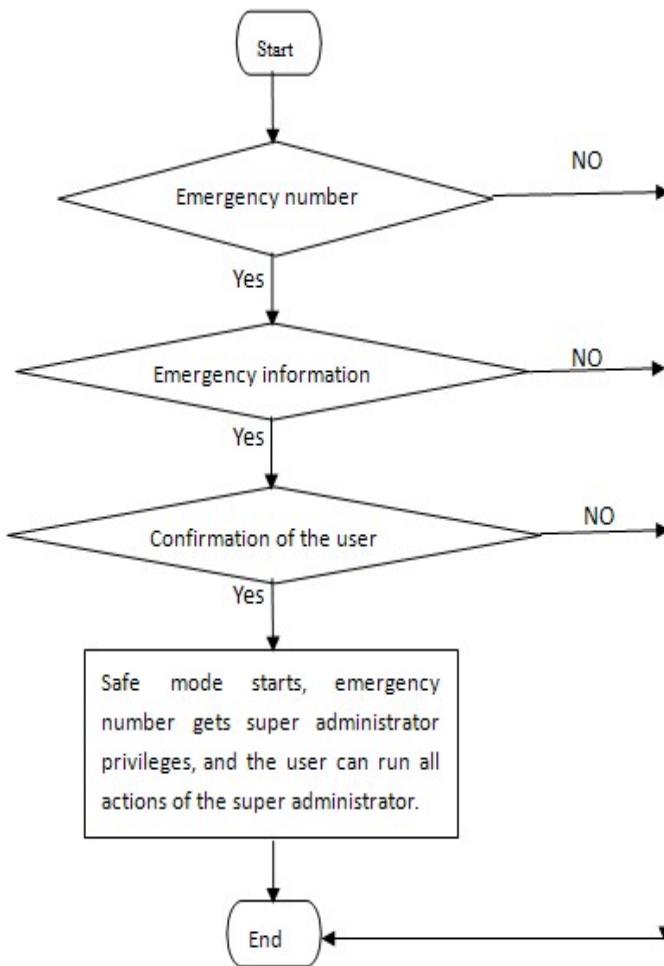
First step: send the password of safe mode starting by the emergency security number;  
Second step: when the mobile terminal sends a confirmation message, the user may reply “Yes” and start safe mode, or reply “No” and cancel safe mode.

After starting safe mode, the emergency security number will become the only super administrator. The user can obtain different data and effect by sending a different message. Reply “1”, the user can obtain all contact information of the current card; Reply “2”, the user will sent the custom information to all people of the phone card; Reply “3”, the user can turn on the camera and take pictures for the man who is using the mobile phone.

The phone will automatically to the emergency security number when SIM card is changed. If the user changes a card, instead of losing the mobile phone, software systems will regard the card as a useful card after the user confirms.

When the user becomes the super administrator, the user sends “stop” through the emergency number, the software can run back to the normal state and close the emergency mode.

The flow chart is shown in figure 3.



**Fig. 3.** Flow chart

In addition, we can prevent data loss after the system reinstall of lost phone through backing up the information database to the cloud.

## 5 Coordination among Three Parts

The paper proposes the solution about the security issues of WSN, home gateway and mobile terminal. But the smart home based IOT makes sure the safety of the three layers, we must firstly ensure their own safety. It is more important to ensure the home environment safe through making their coordination. It may cause the entire information leakage because of the flaws of any parts. In the sensor network and mobile

environment, we can install malicious testing on the intelligent terminal and ensure data privacy and the security of communications data. Thus, according to different scenario, we can ensure the security of communications data by using different encryption methods between the sensor network and the gateway or between the control gateways, using the random key pre-allocation model on key distribution, using digital signatures in the process of information transmission. The Sensor-Gateway-Terminal security mechanism provides a security door for our home environment and ensures the information transmission and privacy on the greatest degree.

## 6 Prospect

The smart home based on IOT is generated on the basis of technology information development and gradual improvement of the living conditions, and is an inevitable trend of development on this new era. The most crucial issue is that protecting the safety about the further development of the home system. The studying on the smart home based on IOT just starts in china currently, but it must be a major research direction in the future.

## References

1. Lü, L., Luo, J.: Smart Home and Its Development Trend. *J. Computer and Modernization* 147, 18–20 (2007)
2. Li, Z.-S.: The Internet Of Things On Security Research. *J. Netinfo Security* 04, 75–77 (2011)
3. Aixia, W.: Intelligent Home Network and Its Security Research. *ShangHai Jiao Tong University* 06 (2008)
4. Erlingsson, Ú., Livshits, B., Xie, Y.: End-to-end Web Application Security. In: Proceedings of the 11th Workshop on Hot Topics in Operating Systems (HotOS XI), San Diego, pp. 1–6 (2007)
5. Jara, A.J., Zamora, A.J., Skarmeta, M.: An Architecture Based on Internet of Things to Support Mobility and Security in Medical Environments. In: Proceedings of the 7th IEEE Consumer Communications and Networking Conference, Las Vegas, pp. 1–5 (2010)
6. Hu, X., Zhang, Y.: Research on Security Mechanism of Nodes Joining in and Quitting from WSN. In: 2011 Third Pacific-Asia Conference on Circuits, Communications and System (PACCS), Wuhan, pp. 1–4 (2011)

# Indicator Framework of IOT Industry Growth Based on Analytic Hierarchy Process<sup>\*</sup>

Jiajun Li<sup>1,\*\*</sup> and Nan Ma<sup>2,\*\*\*</sup>

<sup>1</sup> Management School, Northwestern Polytechnical University,  
710129, Xi'an, China

<sup>2</sup> Economic Research Center, Northwestern Polytechnical University,  
710129, Xi'an, China

**Abstract.** This paper discusses the growth factors which impact the Internet of Things (IOT) industry. Using relevant historical data to extract the corresponding factors such as industry scale, marketing demand, technology innovation in different hierarchical levels, and quantifies these factors. Making use of the Analytic Hierarchy Process (AHP), we build the indicator framework of IOT industry growth. The empirical results show that: these factors, namely IOT industry scale, innovation level and technology progress, all promote IOT industry growth relevantly. Furthermore, this paper forecasts the regularity of IOT industry growth. All of this provides the basis and support of IOT industry growth.

**Keywords:** Internet of Things, indicator framework of IOT industry growth, Analytic Hierarchy Process analysis.

## 1 Introduction

Internet of Things grows on the basis of computer and Internet, using RFID and wireless data communication technology to construct a world covering everything. It is the third technology revolution in the field of information following computer and Internet. China has officially listed it as one of the five burgeoning strategic industries. As the latest product of the application in information communication technology, IOT is still in the growth stage. The market of IOT industry is still a fragmented market, whose industrialization, large-scale business model have not been established. The strength of IOT industry is dispersive and do not form aggregates. With standards, the cost, key technologies, and security privacy problems have gradually been effectively resolved, then the uneven development of the industrial chain will reveal. So the lag of innovation in service application and business model is becoming the bottleneck, which restricts the rapid and healthy development of IOT industry.

\* **Supported Project:** Supported by the Key Project of Soft Science of Shaanxi (Grant NO.: 2011KRZ14).

\*\* The main research directions; System of IOT industry, Risk Control and Decision-making of Industry.

\*\*\* The main research directions; Risk Control and Decision-making of Industry.

Recently, more academia and experts pay attention to Internet of Things research, Zunbai Li, Guisheng Wu (2011) used Technology Roadmap, which is one multi-dimensional analysis tool, to explain the strategic focus of IOT industry layout. These strategic applications are intelligent transportation, smart power and intelligent industry. Matthias Kranz, Münche et al (2010) pointed out that the use of IOT and related technology, which innovate the human-computer interaction (HCI). Cite some cases such as capacitive touch device installed in kitchen facilities and clothes, IOT installed insports and entertainment and so on. Hongbo Zhu (2010) proposed a "multi-domain integration of ubiquitous integrated services ", which is the core technology idea of Internet of Things. He also studied the architecture of wisdom integrated services platform and the corresponding experimental theoretical model. Jingjing Gu (2010) used the network (geometric) topology information, then proposed a novel positioning model of IOT, which depicted the global distribution of density information with local information.

Domestic and foreign scholars on the Internet of Things research mainly concentrate on two levels: First, the technical level for IOT technology research. There are many research in technology design and experiment, security risk, the foreground of technical application, ect ; Second, the industry level. There are only a few research in this field, and the most are macroscopic concepts and framework of understanding stage. Studying from the industry system architecture and model, we can see that although there are some scholars have proposed IOT industry system architecture, and initially given the IOT industry positioning model, but the relevant system framework is limited to the lack of main factors, stratification and in-depth support. This paper focuses on the elements of IOT industry growth based on these above analysis, uses factor analysis, the supported conditions and potential conditions to analyze. And strive to find objective basis to support IOT industry growth.

## 2 Influencing Factors of Industry Growth

Internet of Things belongs to new technology, and the scope of its application is gradually expanding. The applications of IOT are mostly used by high-end users, so that the industrialization process of IOT is just beginning. When in the early stage of industry growth, the marketing demand for IOT begins to rise. The indicator framework of IOT industry growth based on the Analytic Hierarchy is helpful for IOT industry analysis. This paper uses the method of factor analysis to build indicators framework.

### 2.1 Influencing Elements of IOT Industry Growth

At present, the academia has less research for influencing elements of IOT industry growth. There is no unified understanding on this field. Since IOT industry belongs to the high-tech field, its development is based on the electronic and communication equipment manufacturing, electronic computers and other high-tech industries. This study tries to find out the influencing elements of IOT industry growth from the high-tech industry, and then makes the relevant empirical research.

The high-tech industry refers to the industry group which use contemporary state-of-the-art technology to produce high-tech products. Compared to traditional industries, the main features of the high-tech industry are as follows: technology, knowledge, and capital highly intensive, high value-added, short product life cycle, driving effect and so on. Ministry of Science and Technology, Ministry of Finance, State Administration of Taxation jointly organized the preparation of the "high-tech product catalog (2006)". This directory defined high-tech industries as these industries: electronic information, modern transportation, aerospace, advanced manufacturing, new materials, etc. Study on the growth of high-tech industry is mainly launched from the micro-level, macro-level and meso-level. Micro-level mostly discusses the resources of growth on high-tech enterprises; the macro-level studies on the power of high-tech economic development, innovation and so on; the meso-level explores the industry, which mainly concentrates on the relationship between high-tech industrial development and economic growth, and the dynamic factors of high technology industry.

This paper learns from the influencing factors of high-tech industry growth, as well as the considerations of IOT industry scale, the technology research, endogenous innovation capacity, investment in science, technology and the research of industry competitiveness. We explain the situation of IOT industry growth by quantity and quality method. The quantity mainly refers to industrial expansion and marketing demand. And the quality mainly refers to the technology innovation. We select some indicators to reflect the scale of IOT industry, such as marketing demand and technology innovation. These indicators are used as the factors of IOT industry growth.

## **2.2 Assumption about the Influencing Elements of IOT Industry**

The paper is based on three influencing elements of IOT industry growth. Combine the basic environment with characteristics of IOT industry. Then learn the the idea from high-tech industry growth research, and assume the indicators which have effect on IOT industry growth. According to the systematicanalysis principle and the comparability principle, as well as the AHP, we discuss the first-level indicators and the second-level indicators ( $X_i$ ,  $i = 1, 2, 3, \dots, 9$ ), which are used to establish the framework of IOT industry growth including multi-layer influencing indicators (refer to table 1).

This paper selects nine indicators from industry scale, marketing demand and technology innovation. These indicators are used as measure indicators of IOT industry growth framework. The industry scale reflects industry growth. So we can see the state of IOT industry directly. Select enterprise number and employee number, which reflect the scale of inputs. And select gross output value, which reflects the scale of outputs; marketing demand could reflect the growth space, mainly select RFID marketing scale, high-tech products import value and high-tech products export value to measure; technology innovation could reflect the vitality and tension of industry growth, mainly select science and technology institutions, developed projects and R&D intensity, which reflect the innovation of inputs. R&D intensity is the ratio of R&D (internal expenditures) to prime operating revenue.

**Table 1.** Design about the Multi-layer Influencing Elements of IOT Industry

Research Subject	the First-level Indicators	the Second-level Indicators	Emblem
IOT industry growth	Industry scale	Enterprise number (NO.)	X <sub>1</sub>
		Employees number (person)	X <sub>2</sub>
		Gross output value (one hundred million RMB)	X <sub>3</sub>
	Marketing demand	RFID marketing scale (one hundred million RMB)	X <sub>4</sub>
		High-tech products import value (ten thousand dollars)	X <sub>5</sub>
		High-tech products export value (ten thousand dollars)	X <sub>6</sub>
	Technology innovation	Science and technology institutions (NO.)	X <sub>7</sub>
		Developed projects (NO.)	X <sub>8</sub>
		R&D intensity (%)	X <sub>9</sub>

Note: the first-level indicators refer to the three influencing factors of IOT industry growth and the second-level indicators refer to the subdivision factors of the first-level indicators.

### 3 Construct Indicator Framework of IOT Industry Growth

In this empirical research, in order to explain the characteristics and growth regularity of object more fully and accurately, scholar would consider some related multiple indicators. However, all indicators are a reflection of one subject, it inevitably leads to a large number of overlap. This duplication of information would sometimes deny true characteristics and internal rules. Therefore, we hope that there are fewer variables which reflect more information involved in the quantitative research. Factor analysis is a dimensionality reducing method. This method uses multivariate statistical analysis technique to simplify the data. Though the study of internal dependencies among a number of variables, we could explore the basic structure of observational data. Besides, use a handful of "abstract" variables to represent the basic data structure. This paper uses factor analysis to further quantize every indicator, in order to construct indicator framework IOT industry growth.

Seeing that IOT industrialization is just beginning and the relevant data is very limited. Therefore, we intend to use the related data of electronics and telecommunications equipment manufacturing, one upstream industry of IOT industry to measure its influencing factors. Based on 2002-2010 data, we use the SPSS17.0 software factor process, and factor analysis to sort out the elements of IOT industry growth (Note: data are all from the 2003-2011 Statistical Yearbook of high-tech industry in China, and 2003-2011 Statistical Yearbook).

### 3.1 Standardized Indicators

The values of the indicators in this paper have large gap and the units are not the same. In order to eliminate the error caused by the dimension and magnitude differences and make factor analysis equalize each indicator. First of all, we should standardize the original data. The paper uses Z-Score transformation method. The calculating formula is as follow:

$$Z = \frac{x - \bar{X}}{\sigma_x} \quad (1)$$

Where  $\bar{X}$  refers to the average of  $x$ ,  $\sigma_x$  refers to the standard deviation of  $x$ .  $X_i$  ( $i=1, 2, 3, \dots, 9$ ) refers to variables of factor analysis.  $Z_i$  ( $i=1, 2, 3, \dots, 9$ ) represents standardized  $X_i$  ( $i=1, 2, 3, \dots, 9$ ).

### 3.2 Process of Factor Analysis

Factor analysis determines the significance of all indicators based on a large number of sample data. The basic idea is the dimensionality reduction which means a small number of indicators could describe multiple observed variables. The composite indicator (the main factor) is a linear combination of original variables. Through researching the correlation matrix of internal dependencies, we could integrate the observed variables into several main factors. Then show the correlation between main factors.

This paper uses principal component analysis to solve the eigenvalue of the initial common factor, the variance contribution rate and the cumulative variance contribution rate (see to Table 2). Table 2 shows that after extracting two main factors, the cumulative variance contribution rate reaches to 97.177. Therefore, the two main factors are the key factors to determine IOT industry growth. So choosing the two main factors is more appropriate.

**Table 2.** Total Variance Explained

Eigenvalues Main factor	Initial Eigenvalues		Nonrotatory Eigenvalues		Rotatory Eigenvalues	
	Variance (%)	Cumulative (%)	Variance (%)	Cumulative (%)	Variance (%)	Cumulative (%)
A <sub>1</sub>	85.364	85.364	85.364	85.364	84.324	84.324
A <sub>2</sub>	11.813	97.177	11.813	97.177	12.853	97.177

Data source: The above data are based on the 2003-2011 Statistical Yearbook of high-tech industry in China and the 2003-2011 Statistical Yearbook.

Note: Use A<sub>1</sub>, A<sub>2</sub> to describe the original nine variables based on the principle of principal component extraction.

Table 3 shows the scores of rotated factor. It is the coefficient corresponding to each variable among the two main factors. The table is the factor score matrix, which is the the coefficient of factor-score function calculated by regression algorithm. Besides, we could make the main factor score function.

**Table 3.** Main Factor Score Coefficient Matrix

main factor variables	A <sub>1</sub>	A <sub>2</sub>	main factor variables	A <sub>1</sub>	A <sub>2</sub>	main factor variables	A <sub>1</sub>	A <sub>2</sub>
Z <sub>1</sub>	0.138	0.086	Z <sub>4</sub>	0.133	0.042	Z <sub>7</sub>	0.152	0.221
Z <sub>2</sub>	0.133	0.024	Z <sub>5</sub>	0.116	-0.092	Z <sub>8</sub>	0.114	-0.128
Z <sub>3</sub>	0.126	0.031	Z <sub>6</sub>	0.122	-0.054	Z <sub>9</sub>	0.084	0.918

Note: Weighting the product of the factor score coefficient and the corresponding standardized variables, and then we can get the score of the two main factors.

The relationship between main factors and variables are as follows (F<sub>1</sub>, F<sub>2</sub> are scores of A<sub>1</sub>, A<sub>2</sub> separately) :

$$F_1=0.138Z_1+0.133Z_2+0.126Z_3+\dots+0.084Z_9$$

$$F_2=0.086Z_1+0.024Z_2+0.031Z_3+\dots+0.918Z_9$$

Sum the corresponding the contribution rate proportion of main factor. Then we can get the composite score function. The rate is the variance contribution of each factor divided by the total cumulative variance. The composite score formula is as follow:

$$F_3=\sum_{i=1}^m(v_i/k)*F_i \quad (k=v_1+v_2+v_3+\dots+v_m) \quad (2)$$

F<sub>3</sub> is the composite score, m is the number of main factors, v<sub>i</sub> is the rotatory variance contribution rate, k is the cumulative variance contribution rate. According to the formula 2, we could calculate the composite score of IOT industrial growth indicators, then take corresponding value into the functions as follow:

$$F_3=(F_1*84.324+F_2*12.853)/97.177$$

According to the rotatory variance contribution of each factor, we can calculate the score of each factor:

$$F_3=0.131Z_1+0.119Z_2+0.113Z_3+0.121Z_4+0.088Z_5+0.099Z_6+0.161Z_7+0.082Z_8+0.194Z_9$$

Standardize the coefficient in front of each variable, that is IOT industry growth indicators measuring function:

$$F_3=0.118Z_1+0.107Z_2+0.102Z_3+0.109Z_4+0.081Z_5+0.089Z_6+0.145Z_7+0.074Z_8+0.175Z_9$$

According to the measurement function, we can get the indicator framework of IOT industry growth, see to Table 4:

Table 4 shows that in the first-level indicators, the weight of industry scale, marketing demand and technology innovation is 0.327, 0.279 and 0.394 separately. The total weight of industry scale and technology innovation is more than 0.7. So these two indications give an enormous impetus to IOT industry growth. The second-level indicators weights show that the separately weight of enterprise number, employees number, gross output value, RFID marketing scale, science and technology institutions, and R&D intensity technology is more than 0.1. Seeing from the above correlation coefficient analysis, these indicators have significantly difference with other indicators. These factors, namely industry scale, the level of innovation and technology progress, all have strong correlation with IOT industry growth. It is said that these factors have a strong push to the industry growth.

**Table 4.** The Indicator Framework of IOT Industry Growth

Research Subject	the First-level Indicators	the Second-level Indicators	Weight
IOT industry growth	Industry scale	Enterprise number	0.118
		Employees number	0.107
		Gross output value	0.102
		Total	0.327
	Marketing demand	RFID marketing scale	0.109
		High-tech products import value	0.081
		High-tech products export value	0.089
		Total	0.279
	Technology innovation	Science and technology institutions	0.145
		Developed projects	0.074
		R&D intensity	0.175
		Total	0.394

## 4 Conclusion

The empirical analysis above shows that IOT industry scale, technology innovation (belong to important factors) significantly affect IOT industry. There exists a strong correlation between these factors and IOT industry. The reason is that IOT industry has its own characters, such as diverse technique, extensive system, high technology. The marketing scale does not affect industry growth significantly because IOT industry is still in early growth stages of the industry life cycle. New products gradually win the public's preferences on its own characteristics, so marketing demand begins to rise. However, it is not important enough to be the dominant factor of industry growth. With the market outlook is getting better and better, production costs turn lower, marketing demand will strong relate to IOT industry gradually. This factor would actively promote IOT industry growth.

According to the analysis of IOT industry growth indicator framework, we could explore the regular pattern of IOT industry growth primarily: First, both industrial scale and enterprise clusters are the assurance of IOT industry growth. Built on the

basis of the latest scientific and technological achievements, IOT industry needs industrial scale and enterprise clusters so that it could have a grasp of market information, continuous R&D and knowledge creation. At present, the developing mode of IOT industry, such as industry park, industry zone and industry alliance, is consistent with the regularity of IOT industry; Second, technology innovation is the fundamental path to promote IOT industry growth. On the one hand, technology innovation results in technology progress and the elevation of industry productivity directly. On the other hand, the technology innovation could promote product development, and adjust demand structure. It will help expanding the demand space. Third, government support and the market mechanism improvement should be combined with each other. If rely on the natural evolution of market mechanism in the early stage of IOT industry growth, the pace of development would be bound to a fairly slow speed. So government support is necessary. But IOT industry growth is ultimately a kind of economic and market behavior. Therefore, government support should be combined with market mechanism.

This paper is based on the Analytic Hierarchy Process, uses empirical research on indicator framework of IOT growth. Then extract the key influencing factor of IOT industry growth. Finally, explore the regularity of IOT industry growth. This research obtains a new understanding of IOT industry growth framework. Furthermore, it lays the foundation to explore and build the system of IOT industry growth.

## References

- [1] Kranz, M., Münche, Holleis, P., et al.: Embedded Interaction- Interacting with the Internet of Things. IEEE Computer Society (10), 46–53 (2010)
- [2] James, A., Coope, J.: Database Architecture for the Internet of Things. IETE Technical Review 26(5), 311–312 (2010)
- [3] Li, Z., Wu, G.: The Distribution of IOT Industry Based on Technological Roadmap. Enterprise Economy (6), 10114 (2011)
- [4] Zhu, H., Yang, L., Yu, Q.: Technology Ideas and Application Strategies of IOT. Bulletin of Communications of China 31(11), 2–9 (2010)
- [5] Gu, J., Chen, S., Zhuang, Y.: Positioning Model Based on Wireless Sensor Network Topology of IOT. Bulletin of Computer of China 33(9), 1548–1556 (2010)
- [6] Zhao, Y., Ye, C.: An Empirical Study of Chinese High-tech Industry Growth Stage and Its Conversion. Science of Science and Management of S.&T. 32(5), 92–101 (2011)
- [7] Shu, T., Chen, S., Wang, S., et al.: Collaborative Forecasting Method Based on the Influencing Factor of Supply Chain. Systems Engineering-Theory & Practice 30(8), 1363–1370 (2010)
- [8] Wang, B.: Research of Super-conventional and Leaping Development of IOT Industry Based on the A-U Model. Science Technology Progress and Policy 27(24), 79–82 (2010)
- [9] Luo, Z., Xing, Y.: Exploration on the Development Model of IOT Industry. Macroeconomic Research (12), 24–29 (2010)
- [10] Chen, Y., Zhang, H., Zhang, Y.: Thoughts and Suggestions on Chinese IOT Industry Development. Technology Management Research (20), 103–106 (2010)

# **Research on Architecture of the Internet of Things for Grain Monitoring in Storage**

Baisen Xu, Dexian Zhang, and Weidong Yang

College of Information Science and Engineering, Henan University of Technology,  
Zhengzhou, China  
xiaobao.xugao@gmail.com, zdxzzit@hotmail.com

**Abstract.** Grain is a national strategic resource, and is directly related to national security and social stability. The development of the grain internet of Things is highly valued by the government. The paper discusses the basic concept of internet of things, not only mentions the information function model of IOT for grain monitoring, but also proposes the architecture reference model of the internet of things for grain monitoring which is composed of four layers: perception layer, transport layer, processing layer and application layer. It also analyses the function of each layer, and discusses the key technical problems in future research.

**Keywords:** grain monitoring, architecture, internet of things, EPC.

## **1 Introduction**

The Internet of Things (IOT) is a huge system which connects the physical world and the informational world, and is development trend of the next generation network, and is another wave of the world's information industry. The application of IOT involves health care, air monitoring, industrial control, intelligent transportation, government work and other areas, and the development of IOT has been playing a vital role in improving industry information and intelligence, and enhancing overall strength of the industry. Different governments and other scientific institutions pay more attention to IOT.

The Grain is a strategic resource of the nation. It also is the basic material to ensure people's normal life, and directly related to people's livelihood and social stability. At present, it's very urgent to improve information and intelligence level of grain monitoring and to reduce the loss of grain storage. If we want to implement intelligence of grain monitoring, we need to identify and perceive such factors as grain weight, temperature, humidity, air density and so on. Meanwhile we need to perceive data transportation, data fusion and data processing. All of these just refer a typical IOT system. At present, IOT is in an initial stage, and all kinds of related technologies aren't mature. So the major task in the paper is to analyze and to research the architecture and related key technical problems about the IOT for grain monitoring.

## 2 Overview of IOT

### 2.1 The Concept of IOT

The 1995 publication of Bill Gates's book *The Road Ahead* mentions the embryonic form of IOT: "things-things", but it did not attract many people's attentions. In 1999, The concept of IOT was originally proposed by Auto-ID Labs of MIT. When the international telecommunication union (ITU) officially proposed the IOT concept in "the ITU Internet Report 2005: Internet of Things", it declared the era of IOT to come.

At present, scholars in different fields set off a deeper study on IOT from different perspectives. However, different research conclusions result in different understanding of IOT, and so far there has not been absolute a uniform definition about IOT in the international community. Nowadays, most of people accept such definition as "IOT is a kind of network. Specific to say, the chips, more specifically which are embedded inside the objects of the real world, which are further connected with Internet or communication network based on the given agreement, and they process information exchange with each other, to achieve information sharing, intelligence identification, tracking and monitoring".

### 2.2 The Classification of IOT

The basic characteristics of IOT are summed up as the full perception, reliable transmission, intelligent processing and automatic control. The function of information attainment of IOT includes information identification and information perception, and mainly refers to such technologies as RFID (Radio Frequency Identification) technology and sensor technology. So accordingly, the paper divides IOT into IOT based on RFID technology and IOT based on sensor technology. The IOT based on RFID technology is consisted of RFID tags, readers, Savant system, object named resolution services and physical markup language. It implements the global unique code to things through RFID technology, and realizes tracking, traceability and automation management. The IOT based on sensor includes one or more kinds of sensors, communication network and information processing system. Each sensor has a unique label, so that we can collect real-time data, that realize remote security monitoring and air quality monitoring by sensors.

## 3 Application of IOT in Grain

Just as grain, Internet of things can be applied in all aspects of grain, such as grain purchase, grain storage, grain logistics, and information tracking after grain processing.

Application of the IOT in grain purchase can effectively control the grain circulation market and improve grain administrative management efficiency. First of all, it can accurately take grain sales data. Secondly, it can control the situation about grain purchase and sales of enterprise. In addition, grain track can become easier than before.

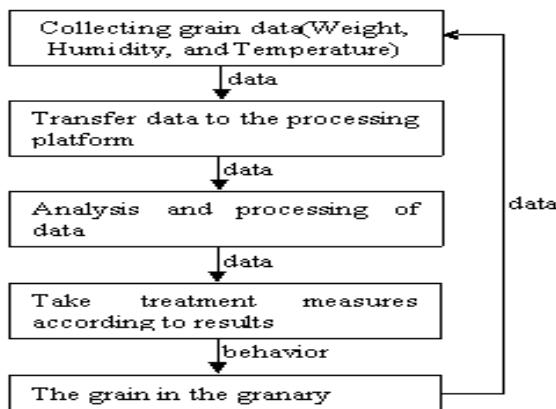
Application of the IOT in grain storage can real-timely control grain data and improve the managerial level of grain storage. Real-time monitoring in grain temperature and humidity by information perception equipment such as sensor, and automatic adjust these conditions in grain storage. And monitoring the change of grain amount in the granary by using pressure sensor can create conditions for reasonable control of the inventory.

Application of the IOT in grain logistics can effectively improve grain transportation efficiency and reduce the losses during the grain transportation. It can not only dynamically monitor the vehicle (or ships) and the grain amount during the transportation, but also realize reasonable dispatch of the vehicle (or ships) on the purpose of ensuring grain quality during the transportation, and realizing the visualization of grain logistics.

Application of the IOT in processed grain sales can effectively guarantee food security. Consumers can easily know the producing area, circulation processing link and product quality problems of bought food by using track functions.

#### 4 Analysis on Architecture of the IOT for Grain Monitoring in Storage

The architecture of the IOT for grain in storage is the prerequisite to build the system of grain monitoring. In order to clearly describe the application of IOT for grain monitoring and to effectively research on the architecture of the IOT for grain monitoring, According to the basic characteristics of IOT and abstract information functional model of the IOT, as shown in figure 1.



**Fig. 1.** Information functional model of IOT for grain monitoring

Collecting data: the grain temperature and humidity is the main factor of affecting grain quality during grain storage, and is the main cause of grain post-harvest losses. The weight of grain plays an important role in rationally allocating inventory for grain

storage authorities. This paper studies on collecting the temperature, humidity, weight and other grain data by a variety of sensors with a unique identification number.

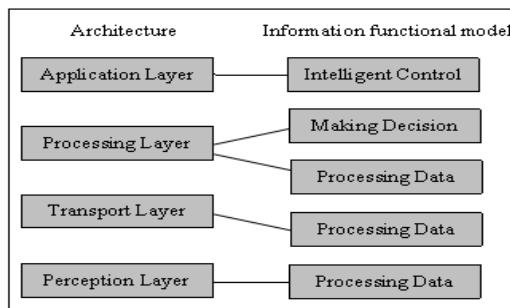
Transferring data: much data collected is stored in the sensor network. The system of the IOT for grain monitoring transfers collected data to data processing platform by communication network or the Internet.

Processing data: the system of the IOT for grain monitoring processes data and extracts effective decision-making information.

Making decision: according to information from processed data, the system of the IOT for grain monitoring makes appropriate decision-making program.

Intelligent control: the system of the IOT for grain monitoring applies established decision-making alternatives to grain or alarm device in the granary, so that we can realize automatic adjustment of storage environment.

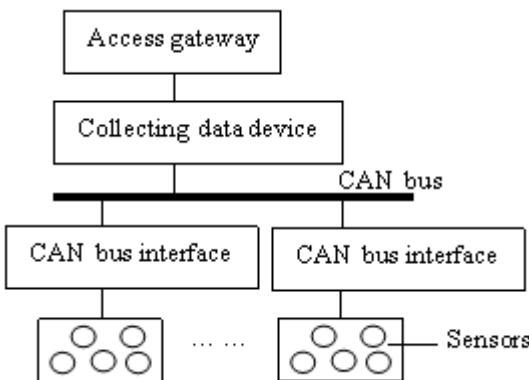
According to information functional model of IOT for grain monitoring, we can divide the architecture of the IOT for grain monitoring into four levels, as perception layer, transport layer, processing layer and application layer. The corresponding relationship between them is shown in figure 2.



**Fig. 2.** Corresponding relationship between system structure and Information functional model

#### 4.1 Perception Layer

During the storage, grain quality can be affected by temperature, humidity and air concentration, and all of these will bring grain damage. Real-timely monitoring these factors is an important means to guarantee the effective storage of the grain. The main task of the perception layer for grain monitoring IOT in grain storage is to collect grain temperature, humidity, air concentration and weight data, and upload the data to access gateway. The main equipment of the perception layer includes pressure sensor, temperature sensor, humidity sensor, air density sensor and corresponding network equipment. The process is to link all kinds of sensors to the CAN bus through the CAN bus interface, collecting the data of all sensors by collecting data device, uploading data information to access gateway, and then transmit the data by access network. The architecture of perception layer is shown in figure 3.



**Fig. 3.** The architecture of the perception layer

#### 4.2 Transport Layer

The transport layer is also known as "the network layer". Its main function is to transmit information to network node, and to further transmit information to processing layer by Internet or mobile communication network. Two networks of the IOT for grain monitoring in grain storage are Internet and mobile communication network. In this layer, we need to study on the access problem among the heterogeneous nets, on network transmission safely problems, and finally ensure the data of grain to correctly transmit grain analysis platform.

#### 4.3 Processing Layer

The main function of the processing layer is to analyze much data about grain situation from all sensors with cloud computing platform or other high-performance computing platform, and provide decision-supporting for the application layer.

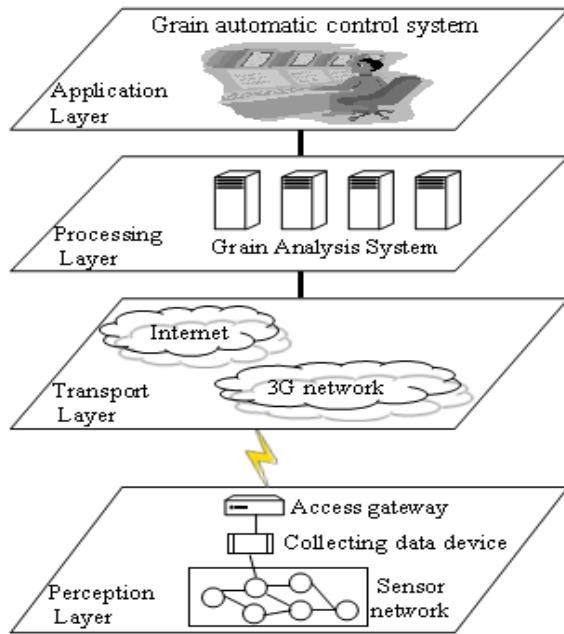
In the layer, we need to research on the effect of the temperature, humidity and air concentration to grain storage quality, and build the relationship model between these factors and grain storage quality, and then calculate the favorable temperature, humidity and air composition for grain storage. According to the conclusion, we need to judge whether the current grain storage conditions are suitable, or not on the result, so that we can provide decision-making support for application layer in time.

#### 4.4 Application Layer

The application layer includes the displaying equipment of grain data interface and the corresponding intelligent control systems. It is the highest layer of the IOT for grain monitoring in grain storage and the purpose of constructing the IOT for grain monitoring in grain storage. According to decision-making information submitted by

the processing layer, the layer real-timely start up automatic control system, which mainly include automatic control temperature system, automatic control humidity system, and automatic regulation of gas composition in the granary and grain quality alarm system and adjust the grain storage conditions. At the same time, automatic assignment inventory is the purpose of constructing the IOT for grain monitoring in grain storage. It can automatically control grain sampling, grain weighing and a variety of valves, and realize the automation of grain in-and-out-of granary, and on the other hand, it also can reasonably adjust grain inventory.

The architecture of the IOT for grain monitoring in storage is shown in figure4.



**Fig. 4.** The architecture of the IOT for grain monitoring in storage

## 5 The Key Problems Which Need to Be Further Studied

### 5.1 The Problems of Grain Data Fusion and Treatment

In order to obtain more accurate and comprehensive perception data information, the system needs to arrange a variety of temperature sensors, humidity sensors, pressure sensors and other sensor nodes, which makes the perception range between two perception nodes overlap. However, this will result in spatial correlation between two adjacent sensors, and cause collected data repeatability and data redundancy. If we don't deal with these redundant data and directly transmit them to gathering nodes, it will cause enormous waste of communication bandwidth. At the same time, a large

number of data uploading will also reduce the communication efficiency. So we need to carry out data treatment such as restructuring, cleaning and fusion before perception data is sent to the gathering nodes.

Due to different of sensors which deployed in the granary, and showed different characteristics, strong space relevance and huge data redundancy among sensors, how to effectively reduce data redundancy and what kind of fusion model and algorithm will be adopted will be our next major research contents.

## 5.2 Sensor Network

The perception layer is a bottom and basic layer of the IOT for grain monitoring in grain storage, which is composed of sensor networks. Sense network comprises with a lot of sensors and communication equipment, and its function is to collect and upload grain data. Therefore, when sensor network fails, it will directly cause decision-making error and application chaos.

Sensor arrangement problem. If the arrangement of these sensors is tight, it will result in a higher degree of data redundancy. On the contrary, it will make spatial correlation poor, and part of the grain data is not collected. So studying on sensor arrangement model will be our next major research content, and ensure time and space relevance between neighbor sensors, and also make collected data have low redundancy.

The networking problem and collaboration-aware problem among neighbor sensors. When the sensor network in which a sensor fails, it may not ensure normal operation of other sensors, and may not ensure that the work of the failed sensor is finished by these adjacent sensors. So we need to research on the networking model of sensor network and collaboration-aware technology among adjacent sensors to ensure normal operation of the IOT system for grain monitoring in grain storage.

Like the Internet, sensor network also faces with network security problem. Network security problem will cause data loss, data tamper and data errors, so we need to implement identity authentication between nodes, to design new encrypt and decrypt algorithm, to establish sensor network security model and to ensure correct transmission of grain situation data.

Sensor arrangement mode, cooperative perception technology, networking model of sensor network and sensor network security model are key technical issues of constructing the IOT for grain monitoring, and all of them are further research contents.

## 6 Conclusion

IOT is a new generation of network on the basis of Internet, and is highly valued by government and research institutions. The application of the IOT in grain, especially in grain monitoring based on grain storage can improve information and intelligence level of grain monitoring, and reduce substantial human, financial resources, and the loss of grain storage. At present, the technical research of grain internet of things is still in its infancy, detailed study in phases is very imperative. This paper focuses on the system structure of IOT for grain monitoring and discusses the basic concept of IOT.

According to the basic characteristics of IOT, this paper doesn't just propose information functional model of the IOT for grain monitoring, but also put forward the 4-layer architecture reference model of the IOT for grain monitoring, such as perception layer, transport layer, processing layer and application layer. The conclusions of the paper possibly have some disputes and shortcomings, but they need to be tested and proved in practice. In the future work, we will discuss the question in detail.

**Acknowledgment.** This work was supported by the Innovation Scientists and Technicians Troop Construction Projects of Henan Province under Grant 094200510009.

## References

1. Atzori, L., Iera, A., Morabito, G.: The Internet of Things: A survey. *Computer Networks* (2010)
2. Shen, S.-B., Fan, Q.-L., Zong, P., Mao, Y.-Q., Huang, W.: Study on the System Structure and Related Technologies for Internet of Things. *Journal of Nanjing University of Posts and Telecommunications* 29(6), 1–11 (2009)
3. Zhu, L., Zhu, Y.-H.: The Usage of Internet of Things in Design of Modern Grain Logistics Monitoring. *Grain Distribution Technology*, 30–33 (2010)
4. Liu, Z.-S., Wei, F., Chai, Y.-T., Shen, X.-S.: Study on the Construction of the Internet of Things in China. *Research and Exploration*, 1–10 (2010)
5. Sun, Q.-B., Liu, J., Li, S., Fan, C.-X., Sun, J.-J.: Internet of things: Summarization on Concepts, Architecture and Key Technology Program. *Journal of Beijing University of Posts and Telecommunications* 33(3), 1–9 (2010)
6. Wang, C.-W.: Analysis on the application prospect of technology Internet of Things in grain logistics. *Cereal & Feed Industry* 8, 12–15 (2010)
7. Wan, Z.-M.: The imagine of building the Internet of Things for grain storage. *State Administration of Grain, Jiangsu* (2010)

# Cloud-Recording Based Intelligent Feedback System of Voice Information

Mengwei Si, Wenwen Du, Yibo Wu, and Jiawei Chen

Shanghai Jiao Tong University, Shanghai, China  
{wssg, chenjiawei}@sjtu.edu.cn, {dewwen, iceworld0324}@126.com

**Abstract.** In recent years, opinion mining becomes a novel and multiuse research topic. This technique has wide and many real-world applications, such as e-commerce, business-intelligence, information monitoring, public-opinion poll, e-learning, newspaper and publication compilation, business management, etc.[1] The goal we develop the Cloud-recording system is to extract information from people's communication through telephones which may be forgotten by accident. Recently, the system can be applied to automate the telephone survey while telephone surveys now are mainly done by humans. With the Cloud-recording system, companies or governments and do the surveys automatically which means telephone surveys can be at a greater amount and a lower cost. In the system, DSR is used to extract features of phones and we translate the phones to text with the help of SPHINX.

**Keywords:** Opinion mining, Opinion extraction, Sentiment analysis, DSR, SPHINX.

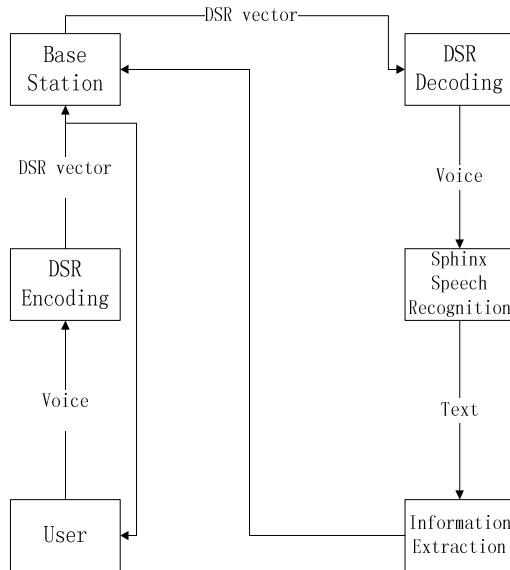
## 1 Introduction

Traditionally, there is no feedback from communication base station to users. People are supposed to take notes for the important information in their phone conversation. Taking notes is not convenient for two reasons. For one thing, one has to take a note book everywhere which is pretty uncomfortable. For another, there are cases in which people cannot take notes such as one is driving a car or swimming in a swimming pool. As a result, a system which can record the telephone call are very important. What's more, people will be more appreciated if the system can extract valuable information automatically. That's why the cloud-recording system is of worth.

The cloud-recording system we have developed enables the analysis of telephone survey to be automatic. Instead of hiring workers, The Company can use a computer to extract the topic and its judgment comments, evaluate whether they are positive or negative and the degree of the comments. All these analyzing process is automatically done. For example, if the respondent is asked that "What do you think of the computer?" and the respondent answers "the computer is very fast", the cloud-recording system will detect that a respondent have assess the theme "computer" and the evaluation is "fast". At the same time, the cloud recording system will find that a respondent have make a positive evaluation become it has made a affirmative assess. Also, strength of the evaluation has been taken into account which indicates that "very fast" is a better evaluation than "fast".

In the system, DSR is used for extract features of phones [2]. There are two merits while using DSR. First, less data are transmitted so the costs for communication are lessened. Secondly, features can be reconstructed to wav form so that we can use SPHINX to translate the phone information into text. The SPHINX is an open source project and with which phones can be translated into text in a relatively high accurate rate [3].

After the phone information has been translated into text, text processing procedure is done. Dependency parsing is accomplished by Stanford Parser. We have design an algorithm to extract theme information and the evaluation of the theme. The system can be expressed in figure 1.



**Fig. 1.** System architecture

## 2 Related Work

### 2.1 Distributed Speech Recognition

A Distributed Speech Recognition (DSR) system overcomes these problems by eliminating the speech channel and instead using an error protected data channel to send a parameterized representation of the speech, which is suitable for recognition.

The value of DSR is that it provides substantial recognition performance advantages compared to a conventional mobile voice channel where both the codec compression and channel errors degrade performance. It also enables new mobile multimodal interfaces by allowing the features to be sent simultaneously to other information on a single mobile data channel such as GPRS [2].

Sphinx is an open-source speech recognition project developed by Carnegie-Mellon University since 1990 [2]. It is also the foundation of current Microsoft speech recognition technique. We use Sphinx to convert voice of wav or raw format into word document, precisely, txt format.

## 2.2 Sphinx

Sphinx has several different versions, among which we choose the one called Pocketsphinx.

An investigation in Sphinx reveals that it demands a user dictionary added by individual user. When analyzing the speech, Sphinx selects the most suitable words that match the voice from the dictionary, and then form a result consists of many separate words with no punctuation. In this term, it is necessary to set a specific application background,

So we can add words which are possible to appear in that condition to the dictionary. After the dictionary is founded and parameters are corrected, Sphinx can run properly under Linux and provide a reliable output.

Besides this, Sphinx also has a training section to improve its performance. It will be studied later carefully since the precision of speech recognition casts a great influence on the subsequent information extraction [3].

## 2.3 Stanford Parser

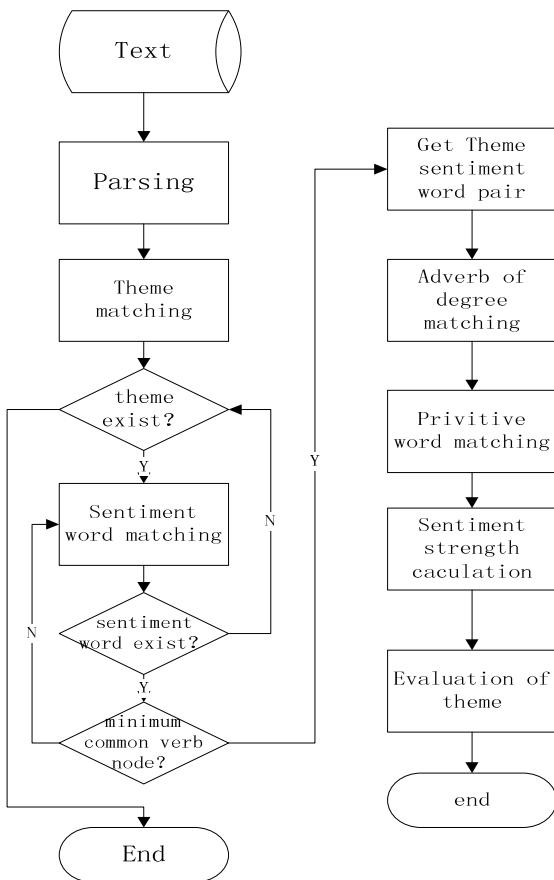
A natural language parser is a program that works out the grammatical structure of sentences, for instance, which groups of words go together (as "phrases") and which words are the subject or object of a verb. Probabilistic parsers use knowledge of language gained from hand-parsed sentences to try to produce the most likely analysis of new sentences. These statistical parsers still make some mistakes, but commonly work rather well [4].

# 3 Structure of the Automatic Telephone Survey System

We have design the application scenario of the telephone survey system as the computer ask the questions automatically and the computer wait for the respondents to answer the questions. The questions are designed by the person who wants to do a survey. And the records of the answers are transmitted to the server part to do phonetic recognition and information extraction. Then the information will be processed by the cloud recording system, the evaluation from the informant would be extracted. Then the person who wants to do the survey will get the result of the survey automatically while in the past the process has to be done by human.

# 4 Algorithm of Evaluation of Theme Extraction

It is designed to recognize the topic and emotional trend of the speech. The algorithm in this part includes the construction of several dictionaries (e.g. dictionary of emotional



**Fig. 2.** Flow chart of algorithm of evaluation of sentimental strength

words, degree adverb words and topic words), matching of thematic and emotional words, dependency parsing and evaluation of the judgment degree.

In the algorithm of evaluation of theme extraction the flaw chart of which is as figure 2, when a text input the system, Stanford parser is applied to parse the text [4]. Then themes and sentimental words will be matched from the dictionary of them. Sentimental words and themes which shared the same minimum common verb node will a theme-sentiment pair. Then the sentimental strength of the sentiment word will be calculated. In order to calculate the strength, adverb of degree and privatives should be taken into account. Then, after considering all the conditions, the algorithm will give the sentimental strength of a theme. Then, sentimental strength of other themes will be calculated in the same way. Hence, we get the evaluation of all the themes in the text.

The following is some explanation of some of the key points.

First, we need to construct them dictionary, sentimental word dictionary, privative dictionary and adverb of degree dictionary. As we have constrained in the area of telephone survey of computer products. The theme dictionary is constructed in the area of computer. For example, computer, keyboard, performance are all in the theme

dictionary. Sentimental word dictionary and adverb of degree dictionary are from Hownet [5]. Some post processes have been done to make the dictionaries suitable to our application. For instance, some of the sentimental words are positive in some areas while negative in other areas. The word “fast” is positive in the sentence “The computer is fast” while “The driver drives too fast” conveying a negative opinion. Thus, we have to redesign the dictionary manually. This process has been done partly. The reasons are twofold. For one thing, the amount of work is too large. For another, it is not necessary to do the work for the influence of small part of sentimental word may not have too much damage to the performance of the system. Privative words dictionary are constructed manually for the reason that the number of privatives is relatively small. As to adverb of degree dictionary, we have to reclassify the words according to our algorithm.

Secondly, the Stanford parser serves as a tool to parse the sentences. For the output of the parser, we can get the pos tagging and parsing of a sentence and the parser provide us a tree of parsing result. We have found that a sentiment word tend to qualify the theme which share the same minimum parent verb node. In the algorithm, we design a method to find the minimum parent verb node. As a result, the algorithm detects the verb and determines the theme-sentiment pair. In the theme-sentiment pair, the sentiment word qualifies the theme. It should be noted that a theme can be qualified by numbers of sentiment words but a sentiment word can only qualify a single theme.

Thirdly, sentiment strength of the evaluation needs to be defined before we determine the sentiment strength of the evaluation of a certain topic [5]. Real number is used to determine the sentiment strength of an evaluation. For example, the sentence “The computer is fast”, sentiment strength of the evaluation is 1 while the sentence “The computer is very fast”, sentiment strength of the evaluation is 2.

Fourthly, adverb of degree enhance the sentimental strength while privatives adverse the sentimental strength. The combine of the two kinds of word sometimes entrance while sometimes weaken the sentimental strength. So we have to classify the adverb of degree carefully to make it distinct. For example, in the sentence “The computer is very fast” the sentiment strength of the evaluation is 2. However, in the sentence “The computer is not very fast” the sentiment strength of the evaluation will never be -2 because the privative waken the sentimental strength [7]. But in fact sometimes, the privative can enhance the sentimental strength such as the sentence “The computer is not a bit fast” [8].

The performance of the algorithm when test in pure text as show in table 1. From the table we can see if the input is pure text the performance of the system is acceptable. However, it is one of the defects of the system that when real phone calls are put into the system, the error rate will increase markedly and we are trying hard to reduce the differences.

**Table 1.** Performance of Algorithm of evaluation of theme extraction

Theme recall rate	Accurate rate of evaluation polarity	F-measure
89.8%	62.9%	76.3%

**Acknowledgments.** This work gets a lot of support from the MediaSoc Lab of Shanghai Jiao Tong University. Here we would like to express our heartfelt thanks to MediaSoc for all its help throughout the work.

## References

1. Yao, T.F., Nie, Q.Y., Li, J.C., et al.: An Opinion Mining System for Chinese Automobile Reviews. In: Proc. of Chinese Information Processing Advanced Progress – Chinese Information Federation 25th Anniversary Conference, pp. 260–281. Press of Tsinghua University (2006)
2. [http://www.etsi.org/WebSite/Technologies/  
DistributedSpeechRecognition.aspx](http://www.etsi.org/WebSite/Technologies/DistributedSpeechRecognition.aspx)
3. <http://cmusphinx.sourceforge.net/>
4. <http://nlp.stanford.edu/software/lex-parser.shtml>
5. Dong, Z., Dong, Q.: Hownet, <http://www.keenage.com>
6. Kim, S.-M., Hovy, E.: Determining the Sentiment of Opinions. In: Proceedings of COLING 2004, the Conference on Computational Linguistics (COLING 2004), Geneva, Switzerland, pp. 1367–1373 (2004)
7. Yao, T., Lou, D., Fang, X.: Polarity Distinction for Chinese Sentiment Words ICCC 2007 paper (IC240)
8. Si, M.W., et al.: Advanced Materials Research 808, 217–218 (2011)

# Shopping System Based on Location Finding Technique of Internet of Things

Bin Wu, Xiao Ling, HongWei Jia, and QiJin Sun

School of Software Engineering,  
Beijing University of Posts and Telecommunications, Beijing, China, 100876  
wubin19896817@163.com

**Abstract.** IOT (Internet of Things) has received some recent attention in information technology research. One reason is that IOT provides a platform that can improve the economic effectiveness and save cost. The other reason is that IOT provides technical support for the development of global economy. This paper describes an application of IOT in Shopping System, the Shopping System that would give the shoppers some services based on the position search. The Shopping System uses the location algorithm to get the shoppers' position in the indoor environment and is able to produce the shortest route from the shoppers to the merchandise based on AStar Algorithm, with a minimal amount of market map information. In the end, the limitations of the application are also discussed, together with possible directions for extending our future work toward better Shopping System.

**Keywords:** IOT, Shopping System, AStar Algorithm, location algorithm in Indoor Environment.

## 1 Introduction

The Internet of Things is a network of Internet-enabled objects, together with web services that interact with these objects [5]. The term Internet of Things has first been used by Kevin Ashton in 1999. The concept of the Internet of Things has become popular through the Auto-ID Center. Radio-frequency identification (RFID) is often seen as a prerequisite for the Internet of Things. If all objects of daily life were equipped with radio tags, they could be identified and inventoried by computers [1].

The Shopping System, which will be introduced in Section 2, is designed based on the location finding technique of Internet of Things. This system originated from a College Student Innovative Experiment Project in 2010. It has six major services: single merchandise positioning, route optimization for multiple merchandises, emergency management, recommending advertisement and recommending store. The main reason for design this system is that the system provides a sufficiently rich and useful platform for shoppers getting significant information in the big market. On one hand, Shopping System is simple enough to permit the shopper to understand the function, and only has a small code size. On the other hand, Shopping System provides a opportunity for study of Indoor Location.

This paper mainly describes a core technique implemented in Shopping System. The technique is based on the location algorithm, which can get the shoppers position by the location nodes. In addition, one aim of our work described below is to explore the feasibility and present practicability of Shopping System.

The outline of the paper is as follows. In the next section, we describe the Shopping System. In the Section 3, we introduce the location algorithm in Indoor Environment. Section 4 presents the Astar Algorithm and explain how to generate the shortest route in the Shopping System; and the result of some experiments are presented in Section 5 and Section 6. In the end, the conclusion of this paper is represented in Section 7.

## 2 Shopping System

Shopping System is designed for saving shoppers' time and improving the performance of shoppers. It is originated from a College Student Innovative Experiment Project in Beijing University of Posts and Telecommunications in 2010. Although Shopping System is very simplistic in many aspects when compared with modern big location system, it remains very pragmatic for the location of big market.

We design the system based on the framework of SSH (Spring+Struts +Hibernate), partly because it is convenient for the programmer to manage the code, and partly because it can reduce the load of programmer. We save all the shoppers' information in the database of server. Before the shopper use this system, he must register firstly. In addition, all the store information is also stored in the database.



**Fig. 1.** The home page of Shopping System

Shopping System consists of six major functions. The details of Shopping System function can refer to the following.

- Single Merchandise Positioning: when you input the target commodity, the system will give you the shortest route from you to the commodity.
- Route Optimization for Multiple Merchandises: when you input multiple merchandises in the system, it will give you the optimization of route on the map.
- Emergency Access: when an accident takes place, the system will provide shoppers a shortest route to the emergency access.
- Recommending Commodities: the system will give you some recommendation based on your previous behavior.
- Comment on Store: you can give grades for each store.
- Emergency Management: when an accident takes place, the manager can reap the distribution for stream of people.

### **3 Localization Algorithm in Indoor Environment**

The localization problem can shortly be formalized as follow. Consider a set of mobile unit (shoppers or merchandises), each one with a fixed position. The localization problem is finding out the position of these nodes. In this paper, we are working with the common assumption of 2-D localization, since the third dimension usually is not of primary interest in indoor environment. Thus the position of a node is a 2-tuple = (), where and are evaluated with respect to an origin O [6].

This section consists of 2 subsections. Subsection 3.1 talks about the basic definition of Localization Algorithm; while subsection 3.2 presents the Localization Algorithm.

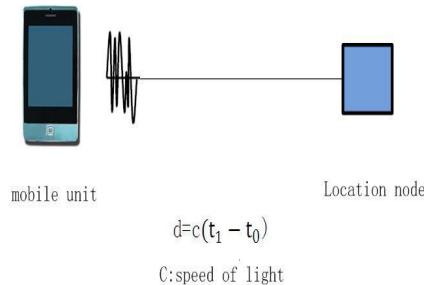
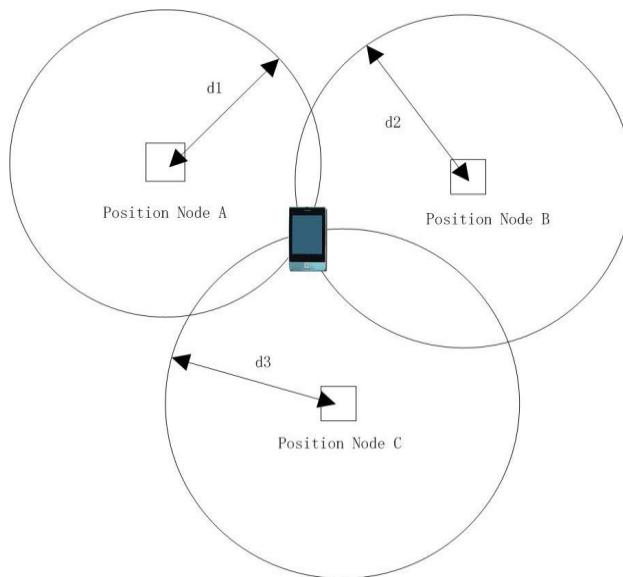
#### **3.1 Basic Definition**

Our location algorithm uses Time-of-Arrival range data, which is measured range data between mobile unit and location node. Time-of-Arrival data is based on the arrival time of radio wave as shown in figure 2. For example, when radio wave is sent time of and received of time, the measured range is calculated as minus in parentheses times c (speed of light) [7].

#### **3.2 Location Algorithm**

In the Shopping System, the shoppers' position can be estimated by trilateration. In the most basic sense trilateration, in two dimensions, is a method using the equations of at least three different circles to find one point where they all intersect [9].

Trilateration estimates the shoppers' position based on the intersection point of the circle, when there are more than three location nodes in the big market [7]. And then the computer can easily get the shoppers' position from the data of location nodes.

**Fig. 2.** Time-of-Arrival based on the arrival time of radio waves**Fig. 3.** Trilateration estimate the shoppers' position

## 4 AStar Algorithm

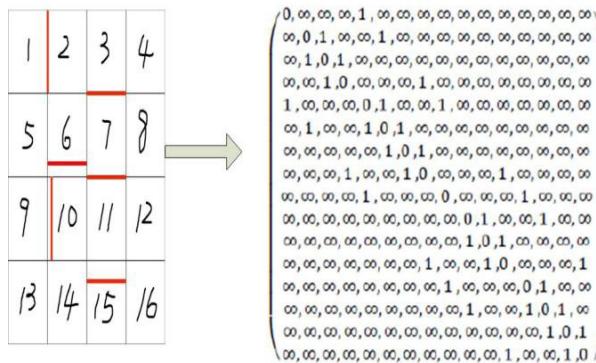
After getting the position information from the location nodes, we choose the AStar Algorithm for calculating the shortest route between shoppers and their merchandises, partly because the efficiency of this algorithm is much higher, and partly because it takes less memory.

For a general market map, we cut the map into a number of squares grids with the same size as small as possible; moreover, we transform the picture information into a two-dimensional array.

#### 4.1 Transform Market Maps into a Two-Dimensional Array

For a general map, it is very difficult for AStar Algorithm to search a shortest route directly. So in Shopping System, we transform maps into a two-dimensional array before we use them. And the implementation process is divided into the following steps: cut the map into a number of squares grids with the same size as small as possible, each grid represents an area. If an obstacle is between two areas, then it represents these two areas are unconnected. For each grid, we consider its four directions, up, down, left and right, judge whether it is connected with its adjacent grids, and define the three rules:

- 1) If two adjacent grids are connected, then the distance between them is 1 unit; otherwise, the distance between them is infinite.
  - 2) The distance between one grid and itself is 0.
  - 3) If there are no obstacles between two grids, then the two grids are connected, else if there are no obstacles on the ligature of midpoints of the two grids, then the two grids are connected, else the two grids are unconnected [2]. A simple example can refer to the following.



**Fig. 4.** A simple example of map transform two-dimensional array [2]

As shown in Fig.4, one market map corresponds to one two-dimensional array. A number represents a grid on the market and a red line represents the wall between two grids.

## 4.2 AStar Algorithm

In computer science, AStar Algorithm is a computer algorithm that is widely used in path finding and graph traversal [3]. This algorithm is first described by Peter Hart, Nils Nilsson and Bertram Raphael. It is an extension of Edsger Dijkstra's 1959 algorithm. As it achieves better performance by using heuristic, AStar Algorithm is widely used. In this paper, as the map is already divided into several small grids, we choose the Manhattan distance, which is much suitable for the grid map, as estimated cost of search. Detail about the procedure of AStar Algorithm can refer to the following.

- Step1: Use nodes to save the information of each grid. Moreover, declare an open set and a close set.
- Step2: Save all the nodes which we never search into the open set, and the nodes which we search before into the close set.
- Step3: For each search, we always get the node that has the lowest distance-plus-cost heuristic function value (usually denoted  $f(x)$ ). The distance-plus-cost heuristic is a sum of two functions:
  - 1) the path-cost function, which is the cost from the starting node to the current node (usually denoted  $g(x)$ ).
  - 2) and an admissible "heuristic estimate" of the distance to the goal (usually denoted  $h(x)$ ) [3].
- Step4: Remove this node from the open set and put it into the close set.
- Step5: For each neighbor of this node, if it is in the close set, search next neighbor node; if it is not in the open set and close set, put this neighbor node into the open set and update the value of  $f(x)$ ,  $g(x)$  and  $h(x)$ ; if it is in the open set and the distance-plus-cost heuristic function value is lower than the previous heuristic function value, then update the open set and save this new route.
- Step 6: Repeat Step 2, Step 3, Step 4 and Step 5 until the following two cases.
  - 1) If the goal node is in the open set, then stop the search and return the shortest route.
  - 2) If the open set is empty, then the route from the shoppers and their goal doesn't exist.

## 5 Path Search Used in Single Merchandise Positioning

In this section, we will present the AStar Algorithm used in the Single Merchandise Position. Firstly, we get the shoppers' position based on the location node in the market. Then, by means of AStar Algorithm, we can calculate the shortest route from the shoppers to their goal merchandise or from the shoppers. In the end, we present the route with red line in the map. The result of route search can refer to Figure 5.

In the Figure 5, we use the red balloons to represent the shopper's position and the merchandise's position. In addition, the red line between these two balloons is the shortest route from the shoppers and the merchandises. The shoppers' position is calculated by Indoor Location Node, and the merchandises' position is stored in the database.

## 6 Path Search Used in the Navigation of Emergency Access

This section will talk about the result of path search in the navigation of emergency access. When the accident takes place, the Shopping System will reap the shoppers' position through Location Algorithm; and then provide the shortest route from the shoppers to the nearest emergency access based on AStar Algorithm. The result of navigation for emergency access can refer to Figure 6.



**Fig. 5.** The shortest route of Single Merchandis Positioning



**Fig. 6.** The shortest route of the Navigation of Emergency Access

In the Figure 6, we use the red balloon to represent the shopper's position. Moreover, the red line between the red balloon and emergency access is the shortest route from the shopper to nearest emergency access. The shopper's position is calculated by the Indoor Location Node, and the emergency access is stored in the database.

## 7 Conclusion

The main contributions of this paper are 1) implementing the AStar Algorithm in the Shopping Systems; 2) presenting a simple way for the Indoor Location. However, a lot of work is required to be done in the future, which involves automatically cutting map into squares grids and Remote Control. Additionally, this Shopping System is never completely finished.

**Acknowledgement.** This paper is supported by College Students Innovative Experiment Project.

## References

1. Internet of Things, Wikipedia, retrieved from,  
[http://en.wikipedia.org/wiki/Internet\\_of\\_things](http://en.wikipedia.org/wiki/Internet_of_things)
2. Ling, X., Wu, B., Sun, Q., Jia, H.: The Improvement of Vertex-exchange Algorithm for the Optimal Hamilton Cycle Problem
3. A\* search algorithm, Wikipedia, retrieved from,  
[http://en.wikipedia.org/wiki/A\\*\\_search\\_algorithm](http://en.wikipedia.org/wiki/A*_search_algorithm)
4. Cormen, T.H., Leiserson, C.E., Rivest, R.L.: Introduction to Algorithm (2001)
5. Introduction of Internet of Things, retrieved from,  
<http://www.internetofthings.net.cn/html/index.php/Index/d/id/59>
6. Pivato, P., Palopoli, L., Petri, D.: Accuracy of RSS-Based Centroid Localization Algorithms in an Indoor Environment. IEEE Transactions on Instrumentation and Measurement (May 2011)
7. Watabe, T., Kamakura, T.: Localization Algorithm in Indoor Environment Based on the ToA Data. In: Circuits and Systems, ISCAS 2005 (2005)
8. Trilateration, Wikipedia, retrieved from,  
<http://en.wikipedia.org/wiki/Trilateration>
9. Trilateration, retrieved from, <http://class.ee.iastate.edu/mmina/ee185/labs/GPS-Activity-2.pdf>

# **Study on the Relationship between Economic Growth and Carbon Emissions Based on Cointegration Theory**

Wenzhou Yan and Yiqing Deng

School of Management, Xi'an Univ. of Arch. & Tech. Xi'an, China  
yan82202896@126.com, kelsey0628@163.com

**Abstract.** In the past ten years, there is an elevating tendency of Xi'an economy development and carbon emissions. Using co-integration analysis method and Eviews6.0 software we proved that there is co-integration relation and bi-directional causality between the Xi'an economy growth and carbon emissions. Then the authors analyzed the pulling effect and the efficiency from the role of the economic growth and industrial structure on carbon emissions by applying the error correction model. Based on the results, we proposed the operation mechanism and policy proposal. It provides theoretical and practical significance on Xi'an low carbon economy development.

**Keywords:** Low-carbon economy, Carbon emissions, Cointegration theory.

## **1 Introduction**

In 2009, the State Council Meeting of energy saving and emissions reduction, presented that the future of economic development in China, should be properly focused on climate change, and based on the characteristic of low carbon emissions develop new economic growth points. For the low carbon economy development, Xi'an has the advantages of new energy development, technology innovation and so on, but the high carbon industry is still dominant, which leads to prominent environmental problems. So actively exploring the Xi'an low carbon economy road is of great significance for resource savings and for an environmentally friendly and sustainable economy. This article will analyze the relationship between carbon emissions and economic development, the influence and the method of developing the low carbon economy.

## **2 Cointegration Theory**

Cointegration theory is the 2003 Nobel laureate in economics, Engle and Granger's research, which is used to describe the long-run equilibrium relationship between two or more non-stationary time series. The economic significance of the cointegration theory is: if two variables have long-term fluctuations in regular, and which is cointegration relationship, there exists the proportion of a long-run equilibrium between the variables.

Definition: If the sequence  $Y_{1t}, Y_{2t}, \dots, Y_{kt}$  are integrated of order 2, there exists a vector  $\beta = (\beta_1, \beta_2, \dots, \beta_k)$ , making  $Z_t = \beta Y_t' \sim I(d-b)$ , in which  $b > 0$ ,  $Y_t = (Y_{1t}, Y_{2t}, \dots, Y_{kt})'$ , then the sequence  $Y_{1t}, Y_{2t}, \dots, Y_{kt}$  is cointegration of order  $(d-b)$  (cointegration), denoted by  $Y_t \sim CI(d-b)$ .

## 2.1 Cointegration Test

Cointegration analysis is the premise that the time series being analyzed are all integrated. So firstly, we need to do a unit root test for the time series. Common methods is Augmented Dickey-Fuller Test (ADF test), that is, through the right side of the regression equation by adding the lagged dependent variable  $y_t$  differential to control the high-end item serial correlation.

When ensuring the non-stationary time series are all  $I(d)$ , you can take the cointegration test, as follows:

If  $k$  - sequence  $Y_{1t}, Y_{2t}, \dots, Y_{kt}$  are order 1 of integration, establish regression equation :

$$y_{1t} = \sum_{i=2}^k \beta_i y_{it} + e_t \xrightarrow{\text{Estimated Residual Sequence}} \hat{e}_t = y_{1t} - \sum_{i=2}^k \hat{\beta}_i y_{it} \quad (1)$$

Where  $i = 1, 2, \dots, k$ . (Refer to cointegration theory definition for the meaning of Symbols sees)

Verify that the residual series  $\hat{e}_t$  is stationary, that is, whether it contains unit roots. If  $\hat{e}_t$  is smooth, the  $k$  variables  $(y_{1t}, y_{2t}, \dots, y_{kt})$  in the regression equation are Cointegration relationships, and the vector is  $(1, -\hat{\beta}_2, \dots, -\hat{\beta}_k)'$ .

## 2.2 Granger Causality Test

On the basis of determining the time series of the Cointegration relationship, in order to determine causal relationships between variables, it is necessary to take the causality test between the two variables. The steps are as follows :

**Using OLS method estimates two regression models, and calculates their respective squared residuals  $RSS_r$  and  $RSS_u$**

Restriction regression model(r):

$$y_t = \sum_{i=1}^s \alpha_i y_{t-i} + e_{1t},$$

Un-restriction regression model (u):

$$y_t = \sum_{i=1}^s \alpha_i y_{t-i} + \sum_{i=1}^k \beta_i x_{t-i} + e_{2t} \quad (2)$$

**Assumes  $H_0$ : adding the lag of the variables x in Ep.2 does not significantly increase the prediction ability of y, so structure F-statistics:**

$$F = \frac{(RSS_r - RSS_u)/k}{RSS_u/(n-s-k)} \sim F(k, n-s-k) \quad (3)$$

**Use statistics F test the null hypothesis  $H_0$ :** for a given level of significance  $\alpha$ , if  $F > F_\alpha$ , reject the original hypothesis that x causes the changes of y; conversely it thinks that x is not the reason for the variation of y.

### 2.3 Error Correction

If in the short term the cointegration variables are not the dynamic structure of equilibrium relationship, error correction model is taken for analysis. Generally applying autoregressive distributed lag (ADL) model and two footwork Engle and Granger (ECM) method. Specifically see the co-integration theory and application [9].

## 3 Xi'an Present Situation of Economic Development and Carbon Emissions

Xi'an's annual rate of GDP growth is steadily on the rise in recent years and is more than 10% GDP value, GDP growth rates for the calendar years and the proportion of the second industrial output value, see table 1.

The Xi'an industry contributes the primary part of carbon emissions. The carbon emissions formula:

$$E_t = \sum_{i=1}^m K_i E_i \quad (4)$$

**Table 1.** Statistics of Xi'an economics of Xi'an economic indicators and carbon emissions data in 1999-2009

Names of Index	GDP (One Hundred Million RMB)	GDP Growth Rates (%)	The Second Industry Scale (%)	The Amount of Carbon Emissions Produced by Coal Consumption(Ten Thousand Tons)	The Amount of Carbon Emissions Produced by Oil Consumption(Ten Thousand Tons)	Total Amount of Carbon Emissions(Ten Thousands Tons)
1999	577.29	10.65	42.15	226.89	10.59	237.49
2000	646.13	12.07	42.89	222.83	21.63	244.46
2001	734.86	11.10	42.58	246.04	35.86	281.91
2002	826.68	12.67	42.77	220.64	47.88	268.53
2003	946.66	14.12	43.03	248.66	62.89	311.56
2004	1102.39	16.09	43.26	386.91	82.22	469.13

**Table 1.** (Continued)

2005	1313.93	14.62	41.14	416.62	100.04	516.67
2006	1538.94	17.11	41.95	462.48	96.88	559.36
2007	1856.63	19.90	42.12	508.80	103.26	612.07
2008	2318.14	14.90	42.34	537.80	114.62	652.43
2009	2724.08	10.65	42.02	586.17	98.38	684.55

Where  $K_i$  is the  $i$  kind of carbon conversion factors for energy. Coal is 0.7476, petroleum is 0.5825 (units: t carbon / Mt);  $E_i$  is the  $i$  kind of energy consumption. According to the primary energy consumption data for 1999-2010, combined with the carbon emissions formula, calculate the amount of Xi'an carbon emissions in 1999-2009, see Table 1.

From Table 1 we can see that the amount of carbon emissions is in a fast growth trend. From 1.68 million tons in 1999 increased to 6.84 million tons in 2009, emissions have increased three times in just 11 years. Particularly after 2003, the rise of carbon emissions is more than 5% per year. Next, this article will use cointegration theory quantitatively analyze the relationship between carbon emissions and economic development and the efficiency in Xi'an.

## 4 Cointegration Analysis about Carbon Emissions and Economic Development in Xi'an

### 4.1 Data Source and Cointegration Test

This paper selects the level of economic development, economic structure and carbon emissions as an analysis object. In order to eliminate heteroscedastic negative effects coming from the time series data, to each variable take natural logarithm. GDP stands for economic development level in years, recorded as G; the second industry scale stands for the economic structure, notes for S; Carbon emissions represented by industrial row carbon, recorded as C. Applying the econometrics software EViews6.0 and ADF test to complete the unit root test, the results see Table 2.

**Table 2.** Results of the unit root test

Variables	Test mode (C,T,K)	ADF test	10% critical value	Conclusion
LNC	(C, 0, 0)	-0.208716	-3.515047	Non-stationary
LNG	(C, 0, 0)	-2.187339	-3.515047	Non-stationary
LNS	(C, 0, 0)	-0.219630	-3.590496	Non-stationary
$\Delta LNC$	(C, T, 1)	-2.252962	-1.598068	Stationary
$\Delta LNG$	(C, T, 1)	-4.487435	-3.590496	Stationary
$\Delta LNS$	(C, T, 2)	-2.250924	-1.597291	Stationary

Note: C—constant, T—time trend, K—a lag order number

Upon examination, after second - order difference of the time series are stationary, so they are second - order integrated sequences of I(2). They meet the cointegration test conditions. Use the method of VAR to take the regression estimates for LNG, LNS and LNC, and get the result:

$$\begin{aligned} \text{LNC} &= -0.544646218646 + 0.872315202528^* \\ \text{LNG} &+ 0.0953799854537^* \text{LNS} + e_t \\ R^2 &= 0.90 ; \quad F\text{-statistic} = 40.18 ; \quad D.W = 0.98 \end{aligned} \quad (5)$$

An Eq.1 r-squared figure (metric goodness-of-fit statistic) is above 0.90, indicating that the sample data regression equation fit the effect very well. Test it through the ADF test and the Engle-Granger cointegration critical value table. The result shows that: the residual sequence  $e_t$  of regression Eq.5 is smooth. It means that Eq.5 is the cointegration relation equation between the LNG, LNS, LNC. This shows carbon emissions and Xi'an's economic development and economic structure have already formed the long-term stable quantity relationship.

## 4.2 LNG, LNS, LNC Granger Causality Test

This article mainly discusses the causal relationships between the Xi'an economic development and carbon emissions, and between the Xi'an economic structure and carbon emissions. So respectively taking the granger causality test on them. Test results are shown in table 3.

**Table 3.** The result of granger Causality test

H <sub>0</sub> hypothesis	F-Statistic	Prob.	Conclusions
LNG does not Granger Cause LNC	0.94127	0.3643	Reject
LNC does not Granger Cause LNG	1.29907	0.2919	Reject
LNS does not Granger Cause LNC	0.99370	0.3520	Reject
LNC does not Granger Cause LNS	6.53368	0.0378	Accept

The results of the granger causality test shows that: " LNG does not Granger Cause LNC " under the significance level of 10 are rejected, description the change of GDP has an impact on carbon emissions; likewise it can be drawn, that the changes of carbon emissions can impact GDP; the adjustment of industrial economic structure, influence carbon emissions, but the change of carbon emissions will not affect economic industrial structure. Therefore, economic development and interaction of carbon emissions are a two-way causal relationship, economic structure and carbon emissions are a one-way causality.

## 4.3 The Error Correction Model of LNG and LNS

Through the OLS estimates, get LNG and LNS (1, 1) order Autoregressive distributed lag models :

$$\begin{aligned} \text{LNC}_t = & 0.3846 * \text{LNC}_{t-1} - 1.68971 + 2.40558 * \text{LNG}_t - 2.07710 * \text{LNG}_{t-1} \\ & + .21037 * \text{LNS}_t + 0.53860 * \text{LNS}_{t-1} \quad (6) \\ R^2 = & 0.99 ; \quad F\text{-statistic} = 18.82019 ; \quad D.W = 1.735872 \end{aligned}$$

R-squared figures in Eq.6 are above 0.99, that means the sample data regression equation fit the effect very well.

$$\text{Make: } \text{ecm}_t-1 = \text{LNCT}-1 + 1.400306799 * \text{LNS}_t-1 - 5.400238137 \text{LNG}_t-1 \quad (7)$$

Deformation of equation (7) :

$$\begin{aligned} \Delta \text{LNC}_t = & -1.6897188 + 2.4055893 * \Delta \text{LNG}_t + \\ & 0.21037479 * \Delta \text{LNS}_t + 0.38463 \text{ecm}_t-1 \quad (8) \end{aligned}$$

Eq.8 is the error correction model of LNG, LNC LNS. Variable coefficient of the variable is the carbon emissions to the change of flexibility: GDP growth of 1% in Xi'an causes the increases of carbon emissions by 2.04%. Error correction for deviation of adjustment is 0.38463. It indicates that when the current issue of the carbon emissions from equilibrium, in order to maintain the co-integration relationship among Xi'an's economic development, carbon emissions and economic structure, the system will be 0.38463 adjustments of the back to balance state.

## 5 Conclusions and Recommendations

This paper used the relevant time series data in 1999-2009 as the sample, using cointegration theory to analyze the relationship among Xi'an's carbon emissions, economic growth and economic structures, and concluded that there is a positive double-action co-integration between Xi'an carbon emissions and economic growth, a positive uni-directional co-integration between Xi'an carbon emissions and the proportion of secondary industry. Because of this relatively long-term stability contact, it makes it relevant that government departments attach importance to promote the rational development of economy, reduce carbon emissions, and promote economic and environmental interaction.

Xi'an is the lifeblood of politics and the economy in Shaanxi Province. It is also the most important industrial city in the west. When active in economic development, it also responds to national "12th five-year-plan" to achieve that the gross carbon dioxide emissions could decrease 18% in this period than in "11th five-year-plan". For that we can establish the following operating system: 1. establish carbon market mechanisms based on carbon trading; 2. establish eco-compensation mechanisms based on carbon tax; 3. establish carbon financial incentive mechanisms; 4. establish government oversight mechanisms. Policy support: 1. Strengthen the publicity of carbon dioxide emissions; 2. strengthen the regulatory role of relevant government departments; 3. using a variety of energy-related taxes, subsidies and other fiscal measures to promote energy conservation; 4. encourage carbon emissions trading on the basis of emission limits. 5. full use of international mechanisms to accelerate the introduction of low-carbon technologies.

**Acknowledgment.** It is supported by the 2009 scientific research project plan of Shaanxi department of education (09JK140).

## References

1. Schmalesee, R., Stoker, T.M., Judson, R.A.: World Carbon Dioxide Emissions: 1950 — 2050. *Review of Economics and Statistics* 80, 15–27 (1998)
2. Galeotti, M.: A Study on Carbon Dioxide Emissions in Developing Countries. In: *Proceedings from the 22ed IAEE Annual International Conference* (1999)
3. Friedl, B., Getzner, M.: Determinates of CO<sub>2</sub> emissions in a small open economy. *Ecological Economics* 159, 133–148 (2003)
4. Wang, H., Tian, P., Jin, P.: The Study of the Relationship Between China's Energy Consumption and Economic Growth Based on Time Varying Parameter Model. *Application of Statistics and Management* 3, 253–258 (2006)
5. Xiao, D., Jiang, Y., Zhao, J.: Co-integration Analysis of the Relationship between Energy and GDP in Shanghai. *Journal of Anhui Agricultural Sciences* 35, 5602–5603 (2007)
6. Xu, C.: Research on the relationship and prediction between China. *Economic Review*, 72 (2010)
7. Liu, J., Yan, W.: Research On the Relationship between Economic Growth in Guangdong Province and Carbon Emissions. *Journal of Guangdong University of Foreign Studies* (21), 19–21 (2010)
8. Ma, W.: Cointegration theory and application, pp. 21–22. D. Nankai University Press (2004)
9. Wang, L., Wan, L.: Econometric theory and application, pp. 181–196. D. He Fei University of Technology Publishing House (2008)
10. Xi'an Bureau, Xi'an Statistical Yearbook. China Statistical Publishing House (2000-2010)

# The Application of Cloud Computing in Smart Grid Status Monitoring

Hongwei Bai<sup>1</sup>, Zhiwei Ma<sup>2</sup>, and Yongli Zhu<sup>1</sup>

<sup>1</sup> School of Control and Computer Engineering

North China Electric Power University

Baoding, China

baihongweinihao@126.com

<sup>2</sup> Department of Electrical Maintenance

Huaneng Qinbei Power Plant

Jiyuan, China

**Abstract.** In the environment of smart grid, the data of condition monitoring will increase greatly. Faced with these massive, distributed, heterogeneous, complex state data, Conventional data storage and management will encounter great difficulties. By the Hadoop cluster technology research and the national grid company's business needs and the actual capabilities of hardware and software, this paper proposes a model based on Hadoop for data processing and conduct simulation experiments to achieve reliable storage and fast parallel processing of the mass data.

**Keywords:** Cloud computing, Smart Grid, distributed storage, parallel processing.

## 1 Introduction

In May 2009, the State Grid Corporation announced the goal of building a unified and strong Smart Grid, the comprehensive and timely grasp of the information of the power grid and integrating automation system for information of the results of the analysis to make the best response reflects the intelligence of the grid. Therefore, accurate, fast, open, shared information system is the basis for the smart grid [1], it requires reliable storage and management of the vast amounts of information, fully taping the potential value of information to enhance the level of analysis and decision support of the Smart Grid [2]. The Information Engineering of State Grid Corporation and China Southern Power Grid Company commonly used conventional solutions, that is, the infrastructure use expensive large-scale server, storage hardware use disk array, database management software use relational database system, the business of tightly coupled classes use software packages, which lead to poor system scalability, higher costs and is difficult to adapt to the requirements of higher reliability and real-time of the state data of the smart grid monitoring .

As an emerging computing model, Cloud computing[3] has high reliability, a huge amount of data processing, flexible, scalable and high capacity utilization benefits, which is becoming a hot research in the field of information, it will bring

opportunities to solve these problems. This paper presents the use of cloud computing technology for smart grid condition monitoring data to achieve the reliable and efficient storage management of all the smart grid business information. It has low cost, high reliability and easy expansion and other advantages which will provide new ideas for the information platform of the smart grid.

## 2 The Introduction of Cloud Computing

Cloud computing is a general term for anything that involves delivering hosted services over the Internet [4]. These services are broadly divided into three categories: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS). The name cloud computing was inspired by the cloud symbol that's often used to represent the Internet in flowcharts and diagrams.

Cloud computing is still in its infancy which is currently used in the Internet [5], commercial and scientific computing. For example, China Mobile has officially identified cloud computing as one of the important direction of the company's strategic and has been building a "big cloud" experiment platform. At present, the in-depth research of cloud computing in the power industry is still rare.

## 3 The Application of Cloud Computing in Smart Grid Status Monitoring

The amount of state data will be increase greatly in the environment of Smart Grid, which will be far beyond the scope of the traditional power grid condition monitoring. That is, it not only covers the equipment of the one system, but also covers secondary system equipment; not only includes real-time online data, but also equipment, basic information, test data, operating data, defect data, inspection records, charged test data and other offline information[6], the amount of data is great and require high reliability and real-time. Faced with these massive, distributed, heterogeneous, complex state data, conventional data storage and management will encounter great difficulties. Based on the characteristics of the smart grid condition monitoring, combining with the open source technology of Cloud computing, this paper proposes the technology architecture of cloud computing platform of the status data of Smart grid data (Figure 1) to meet the needs of smart grid condition monitoring.

Taking into account the status monitoring of smart grid is different from the typical Internet applications, the application of cloud computing on the Internet represented by Google can not be directly applied, this paper is around to open-source Hadoop technology architecture to start. To take full advantage of the idle server resources which belong to the provincial or regional power company, it uses low-cost server clusters. Because the server does not require the same type, it can significantly reduce the cost of construction and use the virtual machine to realize resource virtualization to improve the utilization of the equipment. Of course, although the ratio of low-cost server clusters is high in the performance divided by price, the rate of machine failure is high. This subject use the distributed redundant storage System [7] (HDFS) to store

data to ensure the reliability of data, in other words, it uses high-reliability software to compensate for the high rate of the hardware failure. HDFS is a distributed file system, it has the features of high fault-tolerance and is designed to deploy in a low hardware. HDFS provides high transfer rate to access the application data and fits the applications with large data set (grid monitoring data).

In addition, this subject does not use the traditional relational database, but use Column-based data management [8] (HBase) mode to support the efficient management of large data sets. HBase is a distributed, column-oriented storage system based on the HDFS, which is suitable for the applications of the real-time read and write and random access to large data sets. HBase add nodes by the way of linear fashion from top to bottom to expand, which is not a relational database and does not support SQL. But it can place large and sparse table on the cheap servers skillfully.

Smart grid needs to conduct all kinds of power system computation and application based on state data, for example, condition assessment and fault diagnosis and prediction of the transformer and so on. This subject proposes the state data parallel processing system based on the MapReduce, which can provide high-performance parallel computing and common parallel algorithm development environment for state assessment, diagnosis and prediction. The system mainly consists of two parts: algorithm calls and task management. Scheduling algorithm calls various algorithms which are achieved by the Third-party developers in the form of a plug, for example, fuzzy diagnosis, diagnostic gray system, wavelet analysis, neural networks, and threshold diagnosis and so on. Task Manager is task management, scheduling and monitoring system which is built on the basis of the above-mentioned cloud computing platform based on MapReduce parallel model. MapReduce [9] is a parallel programming model, which divides electricity business logic and the complex details of the distributed computing involved that is concerned by the developers to let the parallel application development of the state assessment, fault diagnosis and prediction mask the underlying implementation details provided by MapReduce programming model.

## 4 The Experiment of Access to Insulator Condition Monitoring Data

### 4.1 The Principle of Design

In the environment of the smart grid, condition monitoring date is wide, panoramic, reliable and mass. The traditional data storage and management of data center obviously can not meet the needs of the smart grid. For example, only leakage current of the insulator monitoring, assuming collecting data once every 10ms, one tower can reach 250 million records [10]. For relational databases, efficiency is extremely low and even intolerable conducting SQL query in a table with 250 million records. Illustrated by the example of inquiring Insulator leakage current (Table 3-1), we prove the correctness of the proposed method. The thinking of design is as follows:

**Table 1.** The Data of Insulator disclosed Current

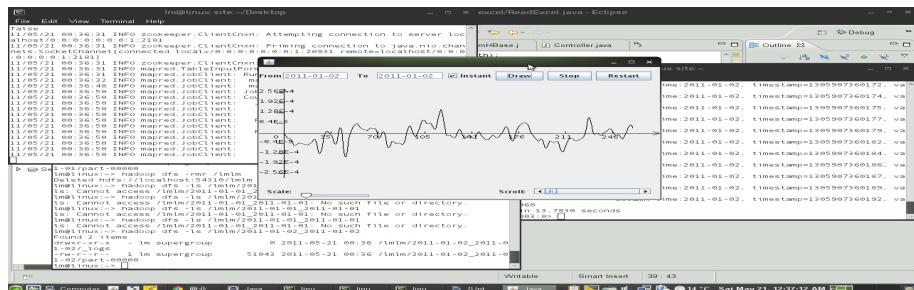
0. 00492	0. 00004714	0. 00491797	0. 00003516
0. 00248	0. 00003614	0. 00003516	0. 00240625
0. 00011	0. 00002896	0. 00240625	0. 00240625
0. 00012	0. 00004522	0. 00003516	0. 00240625
0. 004848	0. 00002885	0. 00003516	0. 00247657
0. 00492	0. 0002547	0. 00247657	0. 00003516
0. 002407	0. 00003458	0. 00247657	0. 00003516

The system is designed to query the needed data stored in HBase and shown as curve, so it only uses the Map Feature in the Map-Reduce. That is, it removes the data which is <time value> <key, value> model and save in the HDFS system to prepare for the following use of display module.

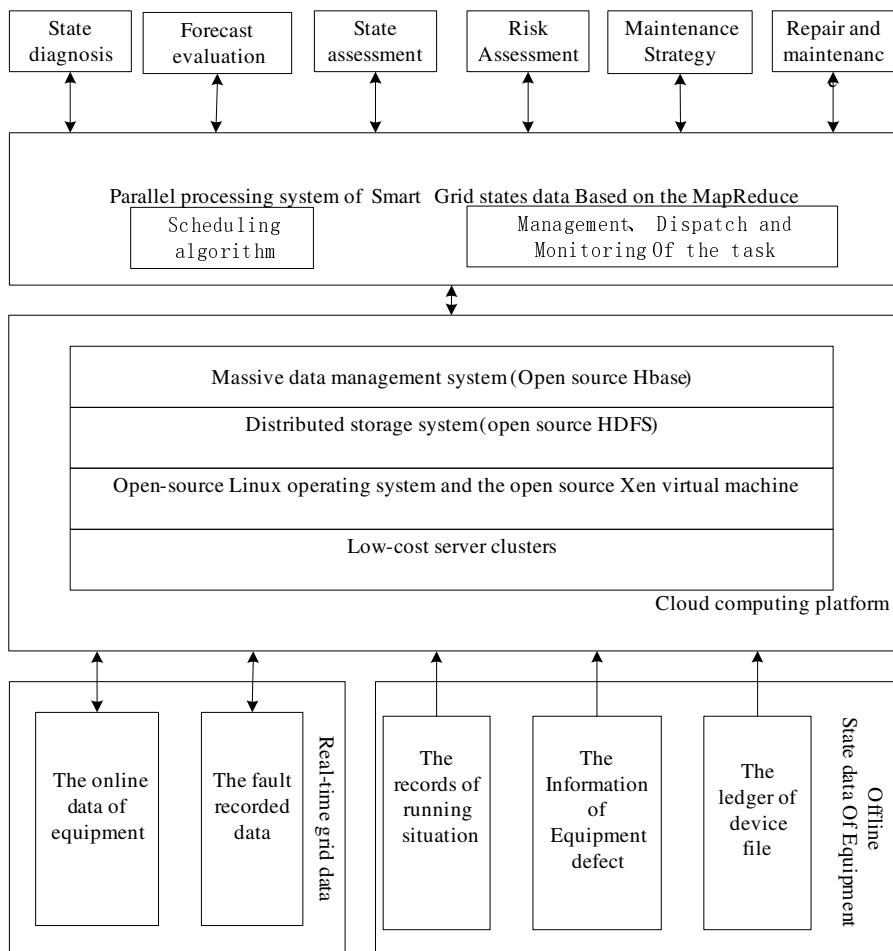
It gets data by the mean of Map-Reduce calling HBase, each line is a processing unit, Map-Reduce will distribute query process to all nodes automatically to improve the efficiency of the query. The Map of query process removing data by the form of <key, value>.That is, the key are string of time-related long values which is calculated according to the system's date + time, the value is the size of the current in this time point. Before the data transport to the Reduce function, the Hadoop system will sort by time from small to big, then the data will be passed by the reduce function and be combined into a single output file.However,Hadoop will take some time to assign tasks. Therefore, this technique is only applicable to relatively low real-time requirements of data mining, processing, etc., but is not applicable to Human-Computer Interaction Design program which has real-time requirements. This subject processes and displays data which is already processed by Map-Reduce that are in the normal OLTP type database cache to improve the response speed. The core code is as follows:

```
public void map(ImmutableBytesWritable row, Result values, Context
context) throws IOException {
    for (byte[] e : values.getRowResult().keySet()) {
        Cell cell = values.getRowResult().get(e);
        if (cell != null && cell.getValue().length > 0) {
            DateFormat formatter = new SimpleDateFormat("yy-MM-dd");
            Long time = -11;
            try {time = formatter.parse(new String(e).substring(5)).getTime()+
Long.parseLong(new String(row.get()))- 10000000;
                context.write(new Text(time.toString()),
new Text(cell.getValue()));} catch (NumberFormatException e1) {
                // TODO Auto-generated catch block
                e1.printStackTrace();
            } catch (ParseException e1) { // TODO Auto-generated catch block
                e1.printStackTrace();
            } catch (InterruptedException e2) {
                // TODO Auto-generated catch block
                e2.printStackTrace();}}}}}
```

The results are as Figure1 :



**Fig. 1.** The experimental results of inquiring Insulator leakage current



**Fig. 2.** The Cloud computing platform of Smart Grid status monitoring

## 4.2 The Analysis of the Experimental Results

The Hadoop system is saving more time with the increasing amount of data, the advantage is more obvious. If our experimental data source is calculated in accordance 2k per record, the amount of 50 million records is only 80G in fact. But the Hadoop platform can clearly show advantages only when dealing with large scale data. Hadoop platform can play powerful data processing capabilities when the deal is the internal state data of the smart grid monitoring.

## 5 Conclusion

This article introduced cloud computing into the smart grid condition monitoring and proposed a new intelligent monitoring platform of the smart grid and data storage and processing methods. It focused on the insulator leakage current as an example of the system architecture, Compared with the traditional method in the retrieval speed, the experiment proved it is more efficient with more data, that is, Hadoop platform can play powerful data processing capabilities.

## References

- [1] Farhangi, H.: The path of the smart grid. *IEEE Power and Energy Magazine* 8(1), 18–28 (2009)
- [2] IEEE SCC21. IEEE P2030 Draft Guide for Smart Grid Interoperability of Energy Technology and Information Technology Operation with the Electric Power System (EPS), and End-Use Applications and Loads (EB/OL),  
[http://grouper.ieee.org/groups/scc21/2030/2030\\_index.html](http://grouper.ieee.org/groups/scc21/2030/2030_index.html)
- [3] Dikaiakos, M.D., Katsaros, D., Mehra, P.: Cloud Computing: Distributed Internet Computing for IT and Scientific Research. *IEEE Internet Computing* 13(5), 10–13 (2009)
- [4] Shi, P., Wang, H., Jiang, J., Lu, K.: Cloud Computing for Research and Implementation of Network Platform. *Computer Engineering and Science* 31(A1), 249–252 (2009)
- [5] Ghemawat, S., Gobioff, H., Leung, S.-T.: The Google File System. In: Proceedings of the 19th ACM Symposium on Operating Systems Principles (2003)
- [6] Zhang, H., Zhu, S., Zhang, Y.: The Research and Practice of Power Transmission Equipment Condition Maintenance Technology System. *Power System Technology* 33(13), 70–73 (2009)
- [7] Borthakur, D.: The Hadoop Distributed File System: Architecture and Design. The Apache Software Foundation (2007)
- [8] Chang, F., Dean, J., Ghemawat, S., et al.: BigTable: a distributed storage system for structured data. *Operating Systems Design and Implementation* (2006)
- [9] Dean, J., Ghemawat, S.: MapReduce: Simplified Data Processing on Large Clusters. *Communications of the ACM* 51(1) (2008)
- [10] Zhu, T., Wang, C., Luo, Z.: Inverter AC side exception of MOA Leakage Current of Guangzhou Converter station. *Insulators and Surge Arresters* (2), 22–29 (2007)

# Design Smart City Based on 3S, Internet of Things, Grid Computing and Cloud Computing Technology<sup>\*</sup>

Min Hu<sup>1,2</sup> and Chang Li<sup>3, \*\*</sup>

<sup>1</sup> High Military Tech Staff Room,

Chinese PLA Defense Information Academy, Wuhan, China

<sup>2</sup> School of Economics and Management, China University of Geosciences, Wuhan, China

<sup>3</sup> College of Urban and Environmental Science, Central China Normal University,  
Wuhan, China

humin@cug.edu.cn

lcshaka@126.com, lichang@mail.ccnu.edu.cn

**Abstract.** With High-technology and society being developed rapidly, there is a trend which is from digital earth to smart earth. Moreover, smart city is one of most important work in smart earth. Therefore, how to realize or design smart city leaves much to be desired. This paper puts forward some strategies and architectures for designing smart city based on geo-spatial information science and technology (GPS, GIS and RS), IT, communication technology, network technology (smart sensor web and ubiquitous sensor network), spatial data mining, high performance computing (grid computing and cloud computing), GPU, artificial intelligence and pattern recognition. The link and elements of smart city as well as applied key technologies in the future are outlined with some typical application instances, which will meet the versatile requirements of smart city service better.

**Keywords:** digital earth, smart earth, geo-spatial information, 3S(GPS,GIS and RS), GPU, grid computing, cloud computing.

## 1 Introduction

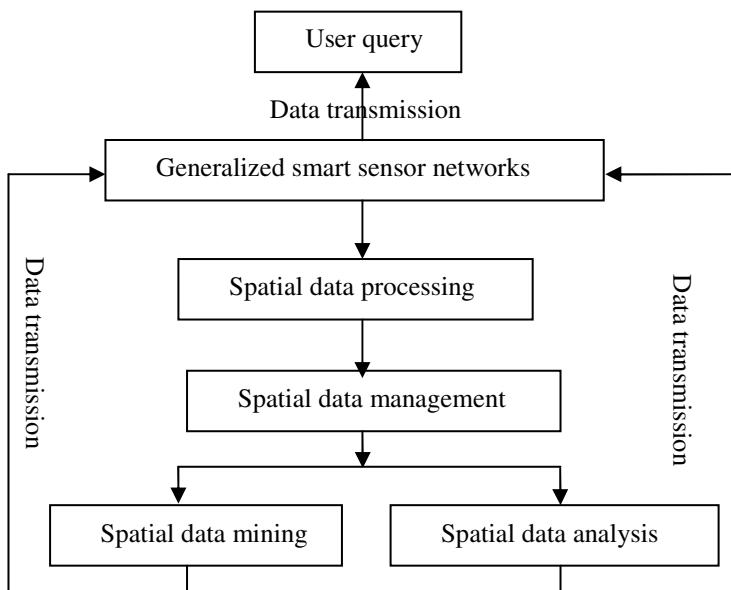
Digital earth is the name given to a visionary concept by former US vice president Al Gore in 1998, describing a virtual representation of the earth that is spatially referenced and interconnected with the world's digital knowledge archives [1]. With Digital Earth improved, a new concept that is Smart Earth is presented by IBM. Smart earth is the first virtual earth website to aggregate renewable energy, sustainable living, new clean technologies and green products/services from around the world into one location [2]. Smart city extended by smart earth is looking for new economic

\* Sponsored by the National Natural Science Foundation of China (41101407, 41001204, 41001260)、Natural Science Foundation of Hubei Province (2010CDZ005)、Self-determined research funds of CCNU from the colleges' basic research and operation of MOE (CCNU10A01001).

\*\* Corresponding author.

growth point after the financial crisis, so it become the world's largest cities' major strategic to propel economic development approaches to change, promote industrial upgrade and revitalize the economy. The concept of the smart city as the next stage in the process of urbanisation has been quite fashionable in the policy arena in recent years, with the aim of drawing a distinction from the terms digital city or intelligent city [3]. In this paper, we focus on the role and key problems of ICT (Information and Communication Technologies) infrastructure, so as to carry out on the role of human capital/education, social and relational capital and environmental interest as important drivers of urban growth.

We design smart city by following main parts: 1) generalized smart sensor networks; 2) data transmission; 3) spatial data processing; 4) spatial data management; 5) spatial data mining; 6) spatial data analysis; 7) user query. The flow chart can be seen in Fig1. In these parts, 3S, internet of things, grid computing, cloud computing, artificial intelligence and pattern recognition can be used to realize smart city.



**Fig. 1.** The flow chart for realizing smart city

Prof. LI Deren stated that “smart earth” can be formed by combining “digital earth” with “internet of things”[4]. Therefore, the organization of the paper is as follows. In section 2, Geo-spatial information science and technology will be discussed for 3D city reconstruction or modeling, which is very important for implementing cyber city. Internet of things technology integrated for digital city is presented in section 3. Section 4 will give the conclusion and state the future work.

## 2 Introduction the Key Technology of Digital City

### 1) Multi-platform and sensor networks

Prof. Tong Qingxi said that Geo-spatial information science and technology is one of the most emerging fields of the progress of science and technology [5]. 3S technology is the key of Geo-spatial information science and technology, which is composed of remote sensing technology (RS), Geography Information Systems (GIS) and Global Positioning System (GPS). And it is the technology of highly combining integrated multi-disciplinary space information's collection, processing, management, analysis, expression, dissemination and application of modern information. Three-dimensional interactive environments offer intuitive and user-friendly ways to view location-based information, such as 3D city models. A 3D city model is usually composed of descriptions of terrain, streets, buildings, other man-made objects and vegetation [6]. In this paper, 3S technology is used for 3D modeling that is the core of cyber city.

The usage of integrated multi-platform, such as aerospace (satellite), aerial (airborne) and ground (vehicle), is a significant trend. Moreover, Professor Gong Peng said that it has the potential of supporting truly integrated ground—space—sky observation of the earth [7]. On the platform aspect. a trend is from emphasizing satellites and airborne platforms to sensor networks on ground that can achieve continuous observation (In this paper, we call these as generalized sensor networks). The development of sensor networks on ground has great importance to remote sensing and geoscientific studies [7]. For example, MMS (Mobile Mapping System) is the process of collecting geospatial data from a mobile vehicle typically fitted with a range of photographic, radar, laser, LiDAR or any number of remote sensing systems. Such systems are composed of an integrated array of time synchronised navigation sensors and imaging sensors mounted on a mobile platform. The primary output from such systems includes GIS data, digital maps, and georeferenced images and video. The development of direct reading georeferencing technologies opened the way for mobile mapping systems. GPS and INS (Inertial Navigation Systems), have allowed rapid and accurate determination of position and attitude of remote sensing equipment, effectively leading to direct mapping of features of interest without the need for complex post-processing of observed data.

### 2) City 3D reconstruction and modeling

Three-dimensional (3D) reconstruction and texture mapping of buildings or other man-made objects are key aspects for 3D city landscapes. In order to realize cyber city, on the ground, we do a lot of researches on 3D-reconstruction for street elevation by means of line matching、solving orientation elements by vanishing point、auto-rectifying and auto-mosaiking large obliquity-angle close-range sequential images、auto-recognizing concavo-convex edge of street elevation [8][9][10][11].

On the sky and space, an effective coarse-to-fine approach for 3D building model generation and texture mapping based on digital photogrammetric techniques is proposed. Three video image sequences, two oblique views of building walls and one vertical view of building roofs, acquired by a digital video camera mounted on a helicopter, are used as input images. Lidar data and a coarse two-dimensional (2D)

digital vector map used for car navigation are also used as information sources. Automatic aerial triangulation (AAT) suitable for a high overlap image sequence is used to give initial values of camera parameters of each image. To obtain accurate image lines, the correspondence between outlines of the building and their line features in the image sequences is determined with a coarse-to-fine strategy. A hybrid point/line bundle adjustment is used to ensure the stability and accuracy of reconstruction. Reconstructed buildings with fine textures superimposed on a digital elevation model (DEM) and ortho-image are realistically visualised. Experimental results show that the proposed approach of 3D city model generation has a promising future in many applications [6].

### **3) Real-time processing—parallel computing**

With the continuous development of sensor technology, to obtain the surface information needs more quickly. Facing diverse data sources and doubling data quantity, many conventional algorithms could not well meet the challenge of the high-speed computing of large-scale data. How to improve efficiency and speed is urgent. Parallel computing is a form of computation in which many calculations are carried out simultaneously, operating on the principle that large problems can often be divided into smaller ones, which are then solved concurrently ("in parallel") [12].

Prof. ZHANG Zuxun presented parallel computing for remote sensing based on blade computer that is called digital photogrammetry grid (DPGrid). Combining the progress of digital photogrammetry hardware and theory, the ideal, that DPW to DPGrid, is educed. The structure and function of a DPGrid are explained, including cluster processing system based on blade computer, fully seam-less mapping system based on network and their features [13][14].

GPU (Graphics Processing Unit) is another method for real-time processing. GPU is a specialized circuit designed to rapidly manipulate and alter memory in such a way so as to accelerate the building of images in a frame buffer intended for output to a display. Modern GPUs are very efficient at manipulating computer graphics, and their highly parallel structure makes them more effective than general-purpose CPUs for algorithms where processing of large blocks of data is done in parallel [15]. The increasing programmability and high performance computational power of GPU present in modern graphics hardware provides great scope for acceleration of photogrammetry and remote sensing algorithms which can be parallelized. With the help of the strong computing ability of CPU + GPU and the parallel computing architecture of CUDA (Compute Unified Device Architecture), CPU + GPU that is used for mass spatial-data real-processing is the direction of future 3S development.

## **3 Internet of Things for City Intelligence**

The Internet of Things refers to uniquely identifiable objects (Things) and their virtual representations in an Internet-like structure. The term Internet of Things has first been used by Kevin Ashton in 1999 [16]. The concept of the Internet of Things has become popular through the Auto-ID Center. Radio-frequency identification (RFID) is often seen as a prerequisite for the Internet of Things. If all objects of daily life were

equipped with radio tags, they could be identified and inventoried by computers [17][18]. However, unique identification of things may be achieved through other means such barcodes or 2D-codes as well. Although the idea is simple, its application is difficult. If all objects in the world were equipped with minuscule identifying devices, daily life on our planet could undergo a transformation [19].

### **1) Sensor web**

The concept of the "sensor web" is a type of sensor network that is especially well suited for environmental monitoring [20][21][22]. The phrase the "sensor web" is also associated with a sensing system which heavily utilizes the World Wide Web. OGC's Sensor Web Enablement (SWE) framework defines a suite of web service interfaces and communication protocols abstracting from the heterogeneity of sensor (network) communication [23].

### **2) Smart city geospatial management**

The GIS operational platform will be the base for managing the infrastructure development components with the systems interoperability for the available city infrastructure related systems. The research will develop Service Oriented Architecture (SOA) in order to geospatially manage the available city infrastructure networks. The concentration will be on the available utility networks in order to develop a comprehensive, common, standardized geospatial data models. The construction operations for the utility networks such as electricity, water, Gas, district cooling, irrigation, sewerage and communication networks, need to be fully monitored on daily basis, in order to utilize the involved huge resources and man power where the SOA will significant value [24]. These resources are allocated only to convey the operational status for the construction and execution sections that used to do the required maintenance. The need for a system that serving the decision makers for following up these activities with a proper geographical representation will definitely reduce the operational cost for the long term.

### **3) Smart city spatial information mining and discovery**

In the future, the Internet of Things must be a non-deterministic and open network in which auto-organized or intelligent entities (Web services, SOA components), virtual objects (avatars) will be interoperable and able to act independently (pursuing their own objectives or shared ones) depending on the context, circumstances or environments. In the future the Internet of Things may be a non-deterministic and open network in which auto-organized or intelligent entities (Web services, SOA components), virtual objects (avatars) will be interoperable and able to act independently (pursuing their own objectives or shared ones) depending on the context, circumstances or environments. Smart city should integrate the physical world with data social, semantic and access networks. Self-organizing networks in a Smart City includes small world overlays with load balancing, semantic gossiping, peer-to-peer reputation-based trust management. There are different forms of self-organization In this Internet of Things, made of billions of parallel and simultaneous events, time will no more be used as a common and linear dimension but will depend on each entity (object, process, information system, etc.). This Internet of Things will be accordingly based on massive parallel IT systems (Parallel computing) such as grid computing and cloud computing which refers to the use and access of multiple server-based computational resources via a digital network (WAN, Internet connection using the World Wide Web, etc.

## 4 Conclusion and Future Work

In an Internet of Things, the precise geographic location of a thing—and also the precise geographic dimensions of a thing—will be critical, so we discuss how to obtain city's 3D spatial information by diverse integrated technologies. Smart city needs that wireless sensor networks, when connected to the Internet, make observation data accessible anywhere and anytime. Therefore parallel computing (grid computing and cloud computing) should be taken into account. However, there are many directions that need to be studied deeply in future, as follows: 1) accurate, reliable and real-time geo-spatial data processing, management, mining and analysis; 2) smart sensor networks (under image condition: CPU + GPU) for internet of things.

## References

- [1] [http://en.wikipedia.org/wiki/Digital\\_Earth](http://en.wikipedia.org/wiki/Digital_Earth)
- [2] <http://sales.smartearth.co/>
- [3] Nicos, K.: Intelligent cities: innovation, knowledge systems and digital spaces. Spon Press, London (2002)
- [4] Li, D., Gong, J., Shao, Z.: From Digital Earth to Smart Earth. Geomatics and Information Science of Wuhan University 35(2), 127–132 (2010)
- [5] Tong, Q.: Earth Observation From Space and Human Demension for Global Change Studies. Advances in Earth Science 20(1), 1–5 (2005)
- [6] Zhang, Y., Zhang, Z., Zhang, J., Wu, J.: 3D Building Modelling with Digital Map, LIDAR Data and Video Image Sequences. Photogrammetric Record 20(111), 285–302 (2005)
- [7] Gong, P.: Some essential questions in remote sensing science and technology. Journal of Remote Sensing 13(1), 1–23 (2009)
- [8] Li, C., Zhang, J., Hu, M.: Auto-reconstructing 3D street elevation for cyber city. Engineering Journal of Wuhan University 42(3), 358–361 (2009)
- [9] Li, C., Li, X., Zhu, A., Li, Q.: Study on auto-reconstructing 3D framework outline of city building. Computer Engineering and Applications 47(8), 4–6 (2011)
- [10] Li, C., Zhou, Y.: 3D Auto-Reconstruction for Street Elevation Based on Line and Plane Feature. In: The 2nd International Conference on Computer and Automation Engineering, Singapore, February 26–28, vol. 1, pp. 460–466 (2010)
- [11] Li, C., Guo, J., Xia, Y.: Image Auto-Mosaic for Large Angle of Obliquity Close-Range Sequential Images to Realize Cyber City. In: 2008 International Symposium on Information Processing, Moscow, May 23–25, pp. 384–388 (2008)
- [12] Almasi, G.S., Gottlieb, A.: Highly Parallel Computing. Benjamin-Cummings publishers, Redwood City (1989)
- [13] Zhang, Z.: From Digital Photogrammetry Workstation (DPW) to Digital Photogrammetry Grid (DPGrid). Geomatics and Information Science of Wuhan University 32, 565–571 (2007)
- [14] Zhang, Z., Zhang, Y., Ke, T., Guo, D.: Photogrammetry for First Response in Wenchuan Earthquake. Photogrammetric Engineering & Remote Sensing 75(5), 510–513 (2009)
- [15] Atkin, D.: Computer Shopper: The Right GPU for You (retrieved May 15, 2007)
- [16] Ashton, K.: That 'Internet of Things' Thing. RFID Journal 22 (Juli 2009); Abgerufen am 8 (April 2011)

- [17] Magrassi, P., Panarella, A., Deighton, N., Johnson, G.: Computers to Acquire Control of the Physical World, Gartner research report T-14-0301 (September 28, 2001)
- [18] Commission of the European Communities. Internet of Things — An action plan for Europe (PDF). COM, 278 final (June 18, 2009)
- [19] Casaleggio Associati The Evolution of Internet of Things (2011)
- [20] Delin, K., Jackson, S.: Sensor Web for In Situ Exploration of Gaseous Biosignatures. In: IEEE Aerospace Conference (2000)
- [21] Delin, K.: Sensor Webs in the Wild. In: Wireless Sensor Networks: A Systems Perspective. Artech House (2005)
- [22] Torres-Martinez, E., Paules, G., Schoeberl, M., Kalb, M.: A Web of Sensors: Enabling the Earth Science Vision. *Acta Astronautica* 53(4-10), 423–428 (2003)
- [23] Botts, M., Percivall, G., Reed, C., Davidson, J.: OGC® Sensor Web Enablement: Overview and High Level Architecture. In: Nittel, S., Labrinidis, A., Stefanidis, A. (eds.) *GSN 2006. LNCS*, vol. 4540, pp. 175–190. Springer, Heidelberg (2008)
- [24] Al-Hader, M., Rodzi, A., Sharif, A.R., Ahmad, N.: SOA of Smart City Geospatial Management. In: 2009 Third UKSim European Symposium on Computer Modeling and Simulation, pp. 6–10 (2009)
- [25] Li, D.: On Generalized and Specialized Spatial Information Grid. *Journal of Remote Sensing* 9(5), 513–520 (2005)
- [26] Li, D., Wang, S., Li, D., Wang, X.: Theories and Technologies of Spatial Data Mining. *Geomatics and Information Science of Wuhan University* 27(3), 221–233 (2002)
- [27] Li, D., Di, K., Li, D., Shi, X.: Mining Association Rules with Linguistic Cloud Models. *Journal of Software* 11(2), 143–158 (2000)

# Analysis and Design of Personalized Recommender System Based on Collaborative Filtering

Jiantao Zhao<sup>1</sup>, Hengwei Zhang<sup>1</sup>, and Yue Lian<sup>2</sup>

<sup>1</sup> Department of Control and Computer Engineering, North China Electric Power University, Beijing, China

<sup>2</sup> Beijing Gaoquan Technology Co., LTD., Beijing, China  
zhaojiantao66@126.com, {zwh0816, lys39290225}@163.com

**Abstract.** Briefly introduces the concept of electronic commerce recommender system, functions and components, proposes the principle of recommender technology, systemically summarizes the various common recommender methods, and simply compares the advantages and disadvantages of these technologies. The collaborative filtering algorithm is emphasized and the advantages of the algorithm are analyzed, combining with the collaborative filtering algorithm based on the user and the item, the article puts forward an improved algorithm, at the same time, the article describes the design process and common problems of the personalized recommender system based on collaborative filtering algorithms. At last, the paper points out that the aspects which still need to be developed and improved, and the future research direction about the recommender system.

**Keywords:** collaborative, recommender, e-commerce, modular, algorithm.

## 1 Introduction

With the vigorous development of e-commerce, the activities of enterprises gradually realizes electronic, e-commerce sites is endless, providing a huge (maybe it is infinite) counter for shoppers, this enables customers to enjoy the happiness and convenient even if they never leave home. But, facing plenty of commodities provided by e-commerce websites, customers can not browse all goods through small computer screen in a short time, another problem is lacking of personal guide, this makes customers face “information overload” problem. Thus, recommender system attracts more and more attention as an important scheme of “information overload” problem.

E-commerce Recommender System is to recommend satisfied target to users according to users' interests, which is also called Personalized Recommender Systems. e-commerce recommender system mainly recommends commodities in the electronic commerce system application, it recommends commodities for users in accordance with their interests. Nowadays, with the rapid development of electronic commerce, the enterprises provide more and more numbers and types of goods in the process of buying and selling, and the various kinds of subsidiary interactive for commodity behavior (ratings, message, etc.) become more and more colorful between the both sides. The sharp increase of the quantity and kinds of commodities brings customs the rich choices, but also brings awkwardness to them when they face the vast ocean of goods, meanwhile, the feedback information of customers also provides much rich marketable

information such as users' preference to merchants, buying habits, etc. E-commerce system itself not only demands applications, but also has the resources, which can be implemented smoothly and widely because of its characteristics.

Collaborative filtering is currently the most widespread and successful technology in the application of electronic commerce recommender system. Collaborative filtering technology applies to algorithms based on the user-based collaborative filtering and the item-based collaborative filtering, the recommender system can enhance the frequency of purchasing goods and improve the satisfaction of users by combining the two algorithms.

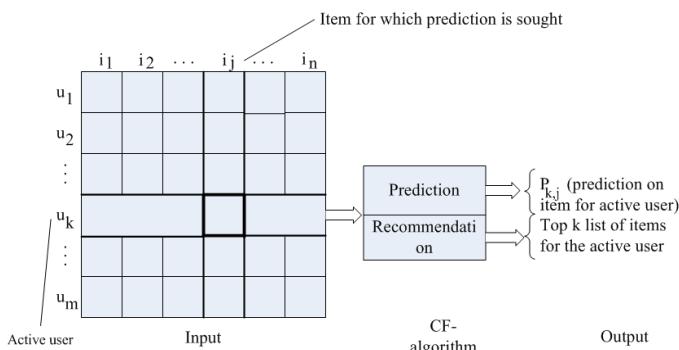
## 2 Collaborative Filtering Technology

### 2.1 Basic Principle

Collaborative filtering technology is a kind of deformation based on storage reasoning, its starting point is no one's interests are isolated, if a group of users value some projects similar scores, then the scores they value on other projects are similar; if a majority of users value some projects similarly, then the current user rates these projects similarly. Collaborative filtering recommender system produces the preference information of target users according to neighbor users' information (the users whose interests are similar to current users).

### 2.2 Collaborative Filtering Processing Steps

The goal of a collaborative filtering algorithm is to suggest new items or to predict the utility of a certain item for a target user based on the users' previous hobbies and the opinions of other like-minded users. Figure 1 shows the schematic diagram of the collaborative filtering process, as the diagram describes, the collaborative filtering processing procedure mainly contains three parts: the algorithm input, algorithm process, and algorithm output.



**Fig. 1.** The collaborative filtering process

First, get algorithm input, namely establish user interest model, this step is mainly to gain the user's interests and transform the information which can be identified by computer. As is shown in figure 1, we construct a  $m \times n$  matrix of the value users rate, the row of the matrix is  $m$  users, and the column is  $n$  items (commodities, etc.). Evaluation value can be displayed such as scores and search keywords, these are given by users, of course, implicit evaluation (click information, collection information, etc.) according to user's behavior is also permitted.

The second step, process input information through algorithms. According to the explicit information and the implicit information, calculate the similarity between users through collaborative filtering, the similarity is the base of recommending commodities to the target users.

The third step, the algorithm output, also namely arise recommender. According to the characteristics of surfing websites, the system timely displays the recommended list, in the actual operation, the general situation is to calculate the values of the items which are related to the target users, then sort the items according to the size of the values. The first  $K$  high items can be recommended to the target users.

### 3 Collaborative Filtering Algorithm

#### 3.1 Classification of the Collaborative Filtering Algorithm

Collaborative filtering has different classification methods according to different levels. According to the difference of the start point, the algorithm contains filtering algorithm based on users-cooperative and filtering algorithm based on items-cooperative.

The collaborative filtering algorithm based on users is also called collaborative filtering based on neighbors or nearest algorithm. This algorithm uses statistical methods to find some neighbor users whose interests are similar with the target user, then the system calculates the hobby degree of the target user according to the neighbor user's interests.

Nearest neighbor search needs a user set  $N = \{n_1, n_2, \dots, n_k\}$  for target user  $u$ , and inside the user set, it must conforms  $sim(u, n_1) > sim(u, n_2) > \dots > sim(u, n_k)$  and  $sim(u, n)$  is the similarity between user  $u$  and user  $n$ ,  $K$  is the threshold, it also called the number in the nearest set.

Estimate the similarity between users mainly rely on related similarity and cosine similarity. Related similarity is measured with person correlation coefficient. If the user  $i$  and the user  $j$  both give the evaluation value to item set, then the similarity degree between user  $i$  and user  $j$   $sim(i, j)$  is:

$$sim(i, j) = \frac{\sum_{a \in I_{ij}} (s_{i,a} - \bar{s}_i)(s_{j,a} - \bar{s}_j)}{\sqrt{\sum_{a \in I_{ij}} (s_{i,a} - \bar{s}_i)^2} * \sqrt{\sum_{a \in I_{ij}} (s_{j,a} - \bar{s}_j)^2}} \quad (1)$$

$S_{i,a}$  represents the score that user i gives to the item a,  $S_{j,a}$  represents the score that user j gives to item a,  $\bar{S}_i$  and  $\bar{S}_j$  separately represent the average evaluation value that user i and user j give to the item. Another method to calculate the similarity is based on vector of cosine angle value. Suppose the evaluating value of user i and user j are respectively  $\vec{i}$  and  $\vec{j}$ , then the correlation similarity between user i and user j is given by

$$\text{sim}(ij) = \cos(\vec{i}, \vec{j}) = \frac{\vec{i} \cdot \vec{j}}{\|\vec{i}\|_2 \times \|\vec{j}\|_2} \quad (2)$$

After gaining the similarity between target user and other users, the system sort the users according to the size of similarity, take the larger former K (threshold) values, then the system can get the corresponding neighborhood of the target user.

Collaborative filtering algorithm based on items inclined to a hypothesis that buyers prefer purchase items that are similar with the items which have been already purchased, the algorithm searches the nearest neighbor set of the target item using statistical techniques, then it predicts the evaluation value according to the evaluation value of the nearest neighbor items which are given by target user, at last, the algorithm selects the top K evaluation value and recommends to the users.

It is similar with collaborative filtering based on users, collaborative filtering algorithm based on items calculates the similarity between the items first, then chooses the neighbor set of the target item. Item similarity calculation first to isolate user i and user j, they both valued the same item, then the algorithm can recommend to the target user some items according the similarity between items (calculation method is similar with the calculation method of similarity between users). As the figure2 shows:

item user \ item	Gongfu pader	Apollo	Big Fish	American Beauty
Zhanghua	4	4	5	5
Liming	2	3	3	3
Wenhe	2	5	3	5
Lihui	4	5	3	

**Fig. 2.** A movie rating matrix

### 3.2 Analysis of the Advantages and Disadvantages of Collaborative Filtering

Collaborative filtering bases on the fact that the users whose interests are close are likely interested in the same thing, users may prefer the commodities that are similar with what they already bought. Compared with traditional filtering method, this method has the following advantages:

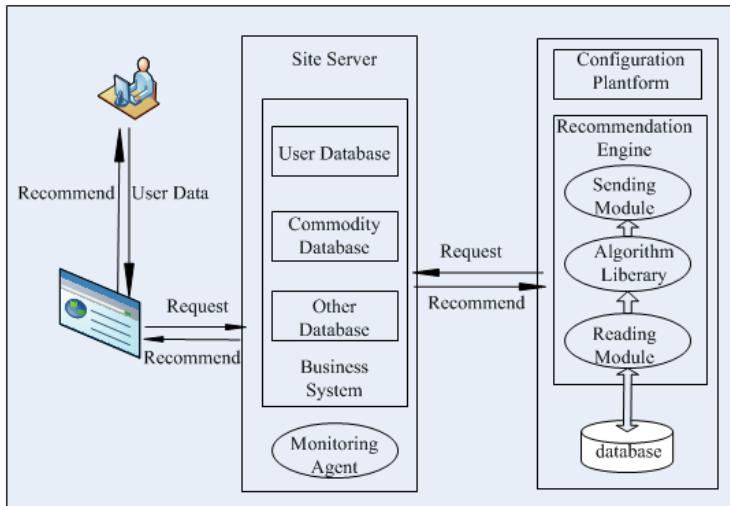
- This method can filter the content that is difficult to analyze automatically by computer, such as music, movies, etc.
- This method makes users share experience with others, analyze content more fully and accurately and can filter items based on some complex concepts such as personal hobbies.
- Have features of finding singularity (the potential preference of users), it can recommend to users some potential but not yet found interests, because the information which is recommended to the user is unexpected.
- It can use the feedback information which is supplied by users themselves, the information given by users is rich and diversity, the method can consider, weight, and comprehensively process the data, this makes the recommender accuracy, validity, and practicability more effective.
- On the other hand, accompanied by the expansion of the structure of network, the complexity of content, the numbers of the products as well as the users, collaborative filtering also faces some challenges such as data sparseness, cold start-up, expansibility, evaluation data model, etc.
- Data sparseness: it is unable to guarantee the accuracy of the nearest users when there is a huge number of commodities but a small number of evaluation values and just a few of items both users valued (the percentage of evaluations is less than 1% relative to the numbers of commodities).
- Cold start-up problem: recommender system cannot guarantee the accuracy of the recommender effect, when some new commodities are added into the system, because no user or less users ever evaluate these commodities, in addition, the fact that the new users give less activity information is another reason.
- The performance of the recommendation algorithm will be sharply down when the scale of the system gets larger, for example, the number of users, the quantity of the commodities increases sharply to another level, in this situation, the system cannot guarantee the recommendation accuracy.

## 4 Design of Recommender System

This section expounds in detail the simple design of recommender system. Recommender system needs to collect rich variety of feedback information from users, as to the input source of recommender data, after getting the recommender data source, the system reads the recommender data, calculates and displays the recommender result set according to the recommendation algorithm warehouse.

### 4.1 The Design of System Architecture

According to the characteristics of the system, the system uses C/S architecture, collaborative filtering system will be isolated from business system, so as to achieve relatively independence and low coupling. The client sends recommender request to the filtering system, collaborative filtering system calculates the data, then the system sets back the recommender result to the business system, and then the result will be displayed on the interface screen. Specific E-commerce Recommender System is as figure3 shows:



**Fig. 3.** The architecture of Recommender System

Before recommending items to the target users, the system first owns a qualified experts library, and presets the categories according to what every expert goods at. This paper points out that the factors of the shopping experts qualification induces two aspects: one is the qualification, namely whether the experts have the ability to provide help for the buyers; another is the influence of the experts in the social networks, namely the experts are willing to help the buyers, and can affect enough buyers about their purchasing behaviors.

## 4.2 Design of Function Module

As the figure 3 shows, Recommender System includes four modules: interface, recommender database, monitoring agency, and recommender engine.

Design of interface: interface (usually called webpage) is the direct communication place between users and the system, users can behave various activities such as value and collect and the commodities, these activities provide the input data for the recommender system. And, after the system makes the recommendation to the target users through recommender engine, the display is shown to the users on the interface. Therefore, the design of interface must be concise and friendly.

Recommender database system: the recommender system needs to design database module due to the historical data of recommender system used to be rich, variety, and huge. In order to improve the portability of the database system, the system is often equipped with a database interface file (similar to the config file of a system), this file is used to help storage engine to link the different kinds of databases. Administrators can also smoothly switch the databases between different kinds of databases, this design can maximize the portability of the data oriented system.

Monitoring agency: the main function of monitoring agency is timely getting the login data and access information, and shows the real-time recommender results to

target users. Monitoring agency monitors in real-time the behavior of the users and timely feedback the information to the recommender engine, after making recommendation for target users, the recommender engine returns the results (usually return HTML format), finally, the monitoring agency gets these HTML pages and show them to the corresponding users.

Recommender engine: recommender engine comprises three modules: recommender reading module, recommendation algorithm library and the send module of the calculate results. Recommender reading module mainly reads the evaluation data given by users. The algorithm of recommender system based on collaboration filtering includes collaboration algorithm based on user and collaboration algorithm based on item, according to the different conditions, and combining various algorithms the system can improve the accuracy of the recommender results. Recommender sending module mainly returns the recommender results through the HTML or XML format to web server, then the server sends the results to the client interface and the results are shown to users.

## 5 Hot Spot and Improving Direction of Research

The traditional personalized recommender systems mainly support the individuation recommenders for single user. It is the trend that the development of multi-user (facing the user group) recommender system which combines with the grid computing. Grid computing is the combination of three kinds of technology: high performance computer, data, and the internet, it can provide a kind of environment which is transparent, sharing, safe and fault tolerance. The grid technology is one of the hot spots which are applied to recommender system. At the same time, integration into the enterprise information system, especially personalized recommender for enterprises, providing decision for product pricing, sales, and management through data analysis have also become the hot spot. And how to establish an effective protection mechanism according to collecting users' behavior and interests which involves the privacy of users is also one of the research directions.

The fundamental of recommender system is the accuracy, and the system can improve the accuracy from three aspects. The first measure is the process of data sources, the system gains diversity data sources and weights the way to process the data, so as to improve the reliability of the input data. The next measure is to improve the algorithm and adopt the improved algorithm, so as to improve the accuracy and effectiveness of the recommender results. The third measure is comprehensive utilization of various recommendation algorithms, combining with the advantages of different algorithms, Eventually reach the purpose of the optimization result.

## 6 Summary

This paper starts from the current situation of the development of recommender system, analyses the basic principle and processing steps of collaborative filtering system which currently is the most extensive application, and the most successful recommender system, concretely analyses the two collaboration algorithms which are

respectively based on the users and the items, and points out the advantages and disadvantages of the two algorithms. The article also designs a electronic commerce recommender system based on the research, introduces the structure model of the system and the functional modules as well as the specific description. At last, this paper introduces and analyses the research spot and the need to further improve the recommender system.

## References

1. Wegner, D.M.: Transactive memory: A contemporary analysis of the group mind. In: Mullen, B., Goethals, G.R. (eds.) *Theories of Group Behavior*, pp. 185–205. Springer, New York (1987)
2. Rokach, L.: Mining manufacturing data using genetic algorithm-based feature set decomposition. *Int. J. Intelligent Systems Technologies and Applications* 4(1), 57–78 (2008)
3. Mobasher, B., Burke, R., Bhaumik, R., Williams, C.: Effective attack models for shilling item-based collaborative filtering system. In: *Proceedings of the 2005 WebKDD Workshop, KDD 2005* (2005)
4. Lam, S.K., Riedl, J.: Shilling recommender systems for fun and profit. In: *Proceedings of the 13th International World Wide Web Conference*, pp. 393–402 (2004)
5. Meththa, B., Hofmann, T.: A survey of attack-resistant collaborative filtering algorithms. *Bulletin of the Technical Committee on Data Engineering* 31(2), 14–22 (2008)
6. Aggarwal, C.C., Wolf, J.L., Wu, K., Yu, P.S.: Hatching Hatches an Egg: A New Graph-theoretic Approach to Collaborative Filtering. In: *Proceeding of the ACM KDD 1999, Conference, San Diego, CA*, pp. 201–212 (1999)
7. Bilsus, D., Pazzani, M.J.: Learning Collaborative Information Filters. In: *Proceedings of ICML 1998*, pp. 46–53 (1998)
8. Peppers, D., Rogers, M.: *The One to One Future: Building Relationships One Customer at a Time*. Bantam Doubleday Dell Publishing (1997)
9. Reichheld, F.R.: Loyalty-Based Management. *Harvard Business School Review* (2), 64–73 (1993)
10. Ungar, L.H., Foster, D.P.: Clustering Methods for Collaborative Filtering. In: *Workshop on Conference on Recommender Systems at the 15th National Conference on Artificial Intelligence* (1998)
11. Oard, D.W., Kim, J.: Implicit Feedback for Recommender Systems. In: *Proceedings of the AAAI Workshop on Recommender Systems*, Madison, Wisconsin (1998)
12. O'Connor, M., Cosley, D., Konstan, J.A., Riedl, J.: PolyLens: A Recommender System for Group of Users. In: *Proceedings of ECSCW 2001*, Bonn, Germany, pp. 199–218 (2001)
13. Avesani, P., Massa, P., Tiella, R.: A trust-enhanced recommender system application: Moleskiing. In: *Proceedings of the 2005 ACM Symposium on Applied Computing*, pp. 1589–1593 (2005)

# The Research of Data Mining Technology of Privacy Preserving in Sharing Platform of Internet of Things

Luyu Chen and Guangwei Ren

College of Computer Science and Information, Guizhou University,  
Guiyang 550025, China  
cluyu0604@sina.com

**Abstract.** The development of the Internet of Things sets off the third waves of the world information industry after the invention and use of computer and Internet. The data mining technology plays a vital role in the development and promotion of Internet of Things, but it causes leakage problem of privacy information at the same time. In the light of the data mining association rules and randomized response method. We propose a new method, suppressible randomized response method (SRRM), and introduce the data mining algorithm of privacy protection based on SRR. Finally, this paper evaluates the privacy of the method.

**Keywords:** Internet of Things, data mining, association rules, privacy preserving, randomized response.

## 1 Introduction

The Internet of Things is a new system that can communicate with the real world. It is also a virtual network including the ubiquitous data perception, the information transmission mainly by wireless and the intelligent information processing through the sharing information platform. Since the prime minister Wen Jiabao proposed “Sensing China”, the Internet of Things was officially classified as the one of the five newest national strategic industries. So the Internet of Things is sweeping over China.

The wide application of the Internet of Things in the production and living must be with more knowledge discovery. In this process, the data mining technology plays a positive role. The data of the sharing platform on the Internet of Things most comes from the data being closely related to people’s life, such as position, environment and habits of work and living. These data are sensitive information for most users, and some malicious users can not access them. So we have to consider the privacy leakage problem in the sharing platform on the Internet of Things.

Without privacy preserving, the Internet of Things may be faced up with a situation that it can not be realized a large-scale commercialization for infringing the privacy of citizens or some collective. Therefore, the studies of how to mine some useful information and protect users’ privacy in the sharing platform on the Internet of Things have a great significance.

## 2 Related Concepts and Work

Data mining is to extract knowledge people are interested in from a large number of data. The association rule is one of them, the definition is as follows:

The following is a formal statement: let  $I = \{i_1, i_2, \dots, i_k\}$  be a set of items. Let  $D$  be a set of transactions, where each transaction  $T$  is a set of items such that  $T \subseteq I$ . We say that a transaction  $T$  contains  $N$ , a set of some items in  $I$ , if  $N \subseteq T$ . An association rule is an implication of the form  $N \Rightarrow Y$ , where  $N \subseteq I$ ,  $Y \subseteq I$ , and  $N \cap Y = \emptyset$ .

The one of the data mining methods in privacy preserving of randomized response method firstly put forward by Warner. The randomized response technology means when we use a random turning needle to handle and involve sensitive problems. If a sensitive problem has two options,  $Y$  and  $N$ , the needle is only seen by respondents and it respectively points to  $Y$  and  $N$  with fixed probability. Finally, according to the probability distribution of needle pointing, investigators get the maximum likelihood estimator of option  $Y$  and  $N$  taking shares respectively in investigation. On the basis of it, this paper proposes a new randomized response method--- suppressible randomized response method (SRRM), which makes the original data be hidden before mining.

## 3 Suppressible Randomized Response Method (SRRM)

In order to express more intuitively, we assume that the data is the data set of market basket. Each commodity is a item with an identified number, customers' each shopping is a transaction, which is expressed by sequence of  $\{0,1\}$ , and the length is the total number of items.

The suppressible randomized response method is firstly exchanging and hiding the original data before the mining of data without the limit of alternatives of randomized parameters. And the specified methods are as follows:

Firstly, giving randomized parameter  $p_a$ ,  $a=1,2,3$ ,  $0 \leq p_a \leq 1$ , and then, set  $f_1=n$ ,  $f_2=1$ ,  $f_3=0$ , in the item  $n$ ,  $n \in \{0,1\}$ , and the  $p_i$  probability selected value of randomized function  $f(n)$  is  $f_j, j=1,2,3$ .

To set the items of total number is  $x$ , the transaction  $N=(n_1, n_2, \dots, n_x)$  which is expressed by a sequence of  $\{0, 1\}$  and the disturbed transaction  $R=(r_1, r_2, \dots, r_x)$  can be calculated through function  $R = F(N)$ , and  $r_i=f(n_i)$ . That is to say, the value of  $r_i$  is  $n_i$  with the probability of  $p_1$ .similarly  $p_2$  equals the probability of  $n_i$ , and  $p_3$  equals the probability of 0.

## 4 The Mining Technology Based on SRRM

By the data transformation and data hiding of SRRM, transaction sets  $D$  can get a forged transaction sets  $D'$ . In the progress of generating the frequent itemsets, the most crucial point is to figure out the support of itemsets. We will introduce how to compute the support of k-itemsets in the following discussion, and then give the mining algorithm based on Apriori algorithm.

#### 4.1 Computing the Support of k-Itemsets

Let  $A=\{i_1, i_2, \dots, i_k\}$  is a k-itemsets. In the case of every item in A which will be handled by the same randomization parameter, we can make use of some optimization strategy to reduce the computing complexity of the itemsets' support. When every item in A uses the same randomization parameter, transaction in D including  $A_i$  will have the same rate with transaction in  $D'$  including  $A_i$  which has been handled by SRRM. That's the reason.

$$l_{ij} = \sum_{t=\max(0, i+j-k)}^{m(i,j)} E_j^t * (P_1 + P_2)^t P_3^{j-t} * \\ E_{k-j}^{i-t} * P_2^{i-t} * (P_1 + P_3)^{k-i-j+t} \quad (1)$$

Regarding to k-itemsets A and transaction T in D, there are  $k+1$  possible value in  $|T \cap A|$ . We take the serial sequence  $E_0, E_1, \dots, E_k$  as the ratio of every transaction in D. For example, as a 3-itemsets, all the transactions in D will be divided into  $\{000\}, \{001, 010, 100\}, \{011, 101, 110\}, \{111\}$ , and  $E_2$  is the ratio of two items in A. Similarly, for the transaction  $T'$  in  $D'$ , there are  $k+1$  possible values in  $|T' \cap A|$  as well. We also take the serial sequence  $E'_0, E'_1, \dots, E'_k$  as the ratio of every transaction in  $D'$ .

$$\text{Then we have } E' = LE, \text{ and: } E' = \begin{bmatrix} E'_0 \\ E'_1 \\ \dots \\ E'_k \end{bmatrix}; \quad E = \begin{bmatrix} E_0 \\ E_1 \\ \dots \\ E_k \end{bmatrix}; \quad L = [l_{ij}] \text{ is a } (k+1) \times (k+1)$$

matrix.  $L_{ij}$  exactly represents that D including  $A_j$  changes into a ratio in  $D'$  including  $A_i$  after it is handled by SRRM.

If L is reversible, let  $L^{-1} = [a_{ij}], E = L^{-1}E'$ , yet  $E_k$  is just the support of k-itemsets which we are computing.

$$E_k = a_{k,0} E'_0 + a_{k,1} E'_1 + \dots + a_{k,k} E'_k \quad (2)$$

Firstly, according to  $D'$  we can get  $E'_j$  and solve  $a_{k,j}$  using L, then the support of k-itemsets A comes out. The time complexity and space complexity of this algorithm is  $O(k)$ .

In addition, we need to notice  $E'_0 + E'_1 + \dots + E'_k = |D'| = N$ , so, there is one item among all the  $E'_j$  can be obtained without any computing. Generally, the value of  $E'_0$  is bigger than anyone else. Because of this reason, we can get the value of  $E'_0$  by way of  $E'_0 = N - (E'_1 + E'_2 + \dots + E'_k)$ .

#### 4.2 The Complete Mining Algorithm

Using the computation formula mentioned above, we can figure out the association rules in which we are interested with the help of various frequent itemsets generation algorithm available. In this paper, using Apriori algorithm, Here is the specific frequent itemsets generative algorithm which has been handled by SRRM.

```

Input: D': The transactional databases handled by SRRM ;
min_sup: The minimum support count threshold.
Output: M: frequent itemsets in D'
        scan D', for each item i∈I count i.count;
        Mi={i∈I | ((i.count/N)-pi)/pi≥min_sup } ;
        for(k=2;Mk,i≠Φ;k++)
            Ek=apriori_gen(Mk-1); // Generate candidate k-itemsets Ek
            for each transaction t∈D'//scan D for counts
            for(j=1; j≤k;j++)
                Et,j=partial_subset(Ek,t,j); // transaction t just
contains candidates k-itemsets of item j
                for each candidate e∈Et,j
                    e.count++;
                for each candidate e∈Ek
                    e.count=ak,0·e.count0+ ak,1·e.count1+ ...+ak,k·e.countk;
                Ek ={e∈Ek | e.count≥min_sup};
        return M=U kMk;

```

## 5 Privacy Assessment

The original intention and ultimate goal of research of privacy protection data mining methods is to do data mining and knowledge discovery, and search the potential, valuable patterns and rules based on the premise that protects privacy data properly. therefore, the level of privacy has become the primary factor when evaluating a kind of method.

According to the Calculation formula of privacy damage coefficient B[4]:

B=P<sub>ratio of real data</sub>×P<sub>probability of real data recognized</sub> +P<sub>ratio of non-real data</sub>×P<sub>probability of non-real data recognized</sub>×P<sub>probability of non-real data reverted</sub>.

Assuming that the ratios of real metadata in all method are the same, computing the privacy damage factor:

Randomization parameter p<sub>1</sub>=p. We take the value p<sub>2</sub>=p<sub>3</sub>=(1-p<sub>1</sub>)/2. In this way, the probability of being 0 and 1 of non-real data will be exactly same, and can't be reverted. Otherwise, non-real data will be possible to be recognized and then reverted. E.g., if we have p<sub>2</sub>=1-p<sub>1</sub>, p<sub>3</sub>=0, then all the data having the value of 0 that has been handled is real data, thereby the protection degree will be reduced greatly. This method which takes the average value of 0 and 1 in practice is not only convenient but also in favor of privacy preserving. In this case,  $B = p_1 \frac{p_1}{p_1 + p_2} = \frac{2p^2}{p+1}$ .

When  $0 < p < \frac{1}{\sqrt{2}}$ , it is a relevant ideal selection range of randomization parameter, and better in privacy.

## 6 Conclusion

In this paper, we proposed a new method of randomized response–SRRM. Then for the data which handled after SRRM, by giving a simple and highly--efficient algorithm to create frequent itemsets, finally to realize a new mining methods of updated associated rule of privacy preserving. We have also analyzed the SRRM way to choose the randomized parameter to strengthen the data's privacy. In brief, the privacy-preserving of sharing on Internet of Things will be one of the hotspots and focal points of Internet of Things industry's development and study, but whether in one of the theoretical level or in the technical level, both of them have many problems which need further investigation and discussion. We hope that more and more effective privacy-preserving data mining algorithms will be proposed and they will play an important role in the application of the Internet of Things widely in the future.

## References

1. Lai, T., Li, W., Liang, H., Zhou, X.: FRASCS:A Framework Supporting ContextSharing. Young Computer Scientists. In: The 9th International Conference for ICYCS 2008, November 18-21, pp. 919–924 (2008)
2. Langheinrich, M.: A Privacy Awareness System for Ubiquitous Computing Environments. In: Borriello, G., Holmquist, L.E. (eds.) UbiComp 2002. LNCS, vol. 2498, pp. 237–245. Springer, Heidelberg (2002)
3. Warner, S.L.: Randomized response: A survey technique for eliminating evasive answer bias. Journal of the American Statistical Association 60, 63–69 (1965)
4. Dong, A.: Privacy-preserving Associatio Rules Mining. Dalian Jiaotong University, DaLian (2007)
5. Agrawal, R., Srikant, R.: Privacy-Preserving data mining. In: Weidong, C., Jeffrey, F. (eds.) Proc. of the ACM SIGMOD Conf. on Management of Data, pp. 439–450. ACM Press, Dallas (2000)

# The SVM and Layered Intrusion Detection System Based on Network Hierarchical

Chao Ju Hu and Jin Wang

School of Control & Computer Engineering  
North China Electric Power University, Baoding Hebei, China  
307471377@qq.com, wangjin2891@163.com

**Abstract.** In this paper, we discussed the general network structure framework, analyzed the most vulnerable attack type for each layer protocol, proposed a design that combined support vector machine (SVM) and layered intrusion detection system (IDS) based on network hierarchy protocol, explained the SVM formula and employed it to analyze.

**Keywords:** intrusion detection system, support vector machine, Network hierarchy, attribute set.

## 1 Introduction

Intrusion detection system (IDS) is a kind of active network security protection system, which is one of the new generation security technology after traditional security technology, e.g. data encryption and firewall etc. Recently, the main technologies in IDS field adopt Bayesian reasoning, artificial neural network, expert system, computer immunology and data mining [1,2]. Some scholars introduced support vector machine (SVM) into intrusion detection and achieved pretty good results. Now the intrusion detection technology based on SVM has become a hot spot in the intrusion detection field.

In order to improve the performance, increase the accuracy and reduce false positives of the intrusion detection system, this paper will use layered network framework and SVM. For each layer of the network, this paper uses SVM to design the most appropriate intrusion detection system and realizes IDS by hierarchical strategy, which can reduce the algorithm training time, improve the efficiency of the algorithm and improve the efficiency of the system.

## 2 Attacks Based on Hierarchical Network Architecture and Network Levels

TCP / IP protocol does not fully comply with the OSI seven layer reference model. The TCP / IP protocol using four layers of hierarchy, each layer call its next level with the network to complete their requirements. These four layers are application layer, transport layer, network layer, and data link layer.

In addition, each layer has its own unique features and each layer also corresponds to a particular attack type of four attacks types.

The application layer corresponding to the attacks: These attacks mainly aim at the network protocol stack application layer attacks, such attacks include: password guessing, buffer overflow, pod attacks, smurf attacks, and so on.

Attacks which transport layer facing: These attacks are targeted at the transport layer network protocol stack attacks, such attacks are: Land attack, Neptune attacks, port scanning, and so on.

Attacks which transport layer facing: These attacks are targeted at the transport layer network protocol stack attacks, such attacks are: Land attack, Neptune attacks, port scanning, and so on.

Attacks which data link layer facing: These attacks are specific to attack link layer network protocol stack, these attacks include: MAC attack, DHCP (Dynamic Host Configuration Protocol) attack, ARP (Address Resolution Protocol) attack, STP and VLAN-Related attacks.

Clearly, the network each layer has its own specific function, so there is a special type of attack based on layer protocol in the internet, so, according to the inherent characteristics of each attack and the means of attack, design a kind of intrusion detection systems distribute on the different levels of network [3,4].

### **3 Support Vector Machine (SVM)**

Support Vector Machine (SVM) proposed by Vapnik in the 1990s is a new type of machine learning algorithm, which built on the strict basis of statistical learning theory, and based on structural risk minimization criterion to obtain the actual risk. Improving effectively the algorithm Generalization ability, it is a better way to solve the small sample of prior learning, nonlinear and high dimension, etc.

Support Vector Machine have a very wide range of applications in IDS systems. Intrusion detection is actually a classification problem which through testing to separate the normal data and abnormal data. But the data of needing disaggregate in IDS is more complex, it often reflects the high-dimensional, small size sample and inseparability [5,6].

Therefore, SVM method is suitable for areas of classification design and the of abnormal findings which is high-dimensional heterogeneous intrusion detection data set and not balance, it will be feasible to apply to the field of intrusion detection.

#### **3.1 The Basic Understanding of Support Vector Machines**

Let an m-dimensional training sample input data  $x_i$  ( $i=1, 2, \dots, M$ ) belonging to Category 1 and Category 2, corresponding to one category  $y_i = 1$ ; the other category  $y_i = -1$ . First discussing the data implementation is divided, in fact not the case (for non-separable non-linear will be introduced after introducing the hard interval support vector machine), then the decision function is:

Formula 1

$$D(x) = W^t x + b$$

Where  $W$  is an  $m$ -dimensional vector,  $b$  is the bias value item. For  $i = 1, 2, \dots, M$ , there be Formula 2.

Formula 2

$$W^t x + b > 0, \text{则 } y_i = 1$$

$$W^t x + b < 0, \text{则 } y_i = -1$$

The training data are linear separable, there is no

$$W^t x + b = 0, \text{ so}$$

Formula 3

$y_i (W^t x + b) > 1, i=1, 2, \dots, M$ , which can be obtained by two Formula .

Hyperplane defined:  $D(x) = W^t x + b = c$ ,

Training data  $x$  and the euclidean distance of classification hyperplane:  
 $|D(x)| / \|W\|$ , The training data are met:

Formula 4

$$\frac{y_k D(x)}{\|W\|} \geq \delta, k = 1, 2, \dots, M$$

Where  $\delta$  is interval.

Hyperplane optimization solution (constraint):

Formula 5

$$\min Q(W) = \frac{1}{2} \|W\|^2$$

$$y_i (W^t x + b) > 1, i=1, 2, \dots, M,$$

The constrained problem into unconstrained problem:

Formula 6

$$Q(W, b, \alpha) = \frac{1}{2} W^t W - \sum_{i=1}^M \alpha_i \{ y_i (W^t x_i + b) - 1 \}$$

$\alpha_i \geq 0$  is non-negative lagrange multipliers.

For  $W$  and  $b$  partial derivatives to get:

Formula 7

$$\frac{\partial Q(W, b, \alpha)}{\partial W} = W - \sum_{i=1}^M \alpha_i y_i x_i = 0$$

$$\frac{\partial Q(W, b, \alpha)}{\partial b} = \sum_{i=1}^M \alpha_i y_i = 0$$

It should also satisfy the Karush-Kuhn-Tucker (KKT) conditions:

Formula 8

$$\alpha_i [y_i (W^T x_i + b) - 1] = 0, \quad i=1, 2, \dots, M$$

The 7 to 6 to be:

Formula 9

$$Q(\alpha) = \sum_{i=1}^M \alpha_i - \frac{1}{2} \sum_{i,j=1}^M \alpha_i \alpha_j y_i y_j x_i^T x_j$$

Therefore, the following conclusions:

If the linearly separable sample set  $\{(x_1, y_1), (x_2, y_2), \dots, (x_M, y_M)\}$ , parameters  $\alpha^*$  is the following quadratic optimization problem.

Formula 10

$$\max Q(\alpha) = \sum_{i=1}^M \alpha_i - \frac{1}{2} \sum_{i,j=1}^M \alpha_i \alpha_j y_i y_j x_i^T x_j$$

Constraints are:  $\sum_{i=1}^M \alpha_i y_i = 0 \quad \alpha_i \geq 0, i=1, 2, \dots, M$

The weight vector  $W^* = \sum_{i=1}^M \alpha_i^* y_i x_i$  determines the optimal hyperplane. At this point the best classification decision function is:

Formula 11

$$D(x) = \sum_{i \in S} \alpha_i y_i x_i^T x + b$$

B does not appear on the dual problem, using the original data can be obtained:

Formula 12

$$b = y_i - W^T x_i$$

Taking the average is

$$b = \frac{1}{|S|} \sum_{i \in S} (y_i - W^T x_i)$$

### 3.2 SVM Feature Space Mapping and Kernel Feature

The support vector machine of hard interval has many shortcomings, because it always tried to produce a consistent hypothesis of No training error. However, when training data are noisy, the feature space generally can not be linear separated. The support vector machine of soft interval can be extended to linear inseparable case.

Non-linear vector function  $g(x) = [g_1(x), g_2(x), \dots, g_l(x)]$ , If the m-dimensional input vector  $x$  will be mapped to the L-dimensional feature space, linear decision function in feature space as follows:  $D(x) = W^T g(x) + b$

Classification decision function is:

$$D(x) = \sum_{i \in S} \alpha_i y_i H(x_i, x) + b$$

$H(x_i, x)$  is Kernel functions, commonly used kernel function is:

Linear kernel function:  $H(x, x') = x^T x'$

Polynomial functions:  $H(x, x') = (x^T x' + 1)^d$

RBF kernel function:  $H(x, x') = \exp(-r \|x - x'\|^2)$

If the SVM kernel function  $H(x, x_i)$  uses the RBF kernel:  $\exp(-g \|x - x_i\|^2)$ , then the discriminant function is:  $f(x) = \text{sgn}\{\sum_{i=1}^M a_i y_i \exp(-g \|x_i - x\|^2 + b)\}$

According to discriminant function, it can be seen that Kernel functions is the decreasing function for the distance of the new data and a sample. That is, when the new data and a sample with a relatively large distances, Without considering the factor's impact, the sample's contribution is relatively small in determining the new data. The coefficient  $\alpha = \alpha * = (\alpha * 1, \dots, \alpha * n)$  has been solved in the training, and no longer change when determining. Thus in the case of higher speed, you can not consider using new data far from the sample. Prediction of new data calculate the distance for new data and samples, keep data close to the sample, and Removed the sample longer distance. With the increasing of the projected cost, when the sample size is large, it is more harm than good. Training samples will be separated by spatial location in the training, The sample data with short distance is divided into a training data class. when prediction, it need only determine the new data proximity with which the data class, then use this sample class data to predict. It not only can greatly improve training speed, since reduction of support vector machines, but also can improve the prediction speed [7,8].

### 4 Description of the Data to Be Detected

The data stream in the network contains a lot of different types of attacks, therefor these data used in the IDS will be classified as follows:

In 1998, Agency WenkeLee and others work in the U.S. Defense Advanced Research Projects for IDS evaluation obtained data. The original data is the connection information which restored on the basis of these data.

The original data is from the data that in 1998 WenkeLee and others, the U.S. Defense Advanced Research Projects Agency (A) for IDS evaluation on the basis of data obtained when the recovery out of the connection information, seven weeks of network flow was included in these data, and there are about five million connection records, including a large number of normal network flow and a variety of attacks, and they have a strong representation. There are four classes of attacks:

DOS \*- denial the service attacks, such as SYN Flood, land attack;

R2L: obtain Remote access, such as: password guessing;

U2R: all kinds of privilege escalation, such as: a variety of local and remote Buffer Over flow attacks;

Probe: a variety of port scanning and vulnerability scanning.

A complete TCP connection session is considered to be a connection record. Each UDP and ICMP packet is also considered to be a connection record. each connection information includes the following four categories of attributes:

- a) Basic set of attributes, such as: duration of the connection, protocol, service, number of bytes sent, bytes received, etc.;
- b) The content attribute set, using domain knowledge to obtain property from the pocket content. Such as: the number of the connecting 'hot' mark, the number of failed login connection, the success of landing;
- c) flow properties set, which is based on time and network traffic-related properties, such property is divided into two sets, one set is SameHost property, that has the protocol behavior, service and some other statistics in the past 2S with the current connection in connection with the same objective connection in the past 2s, another set is the SameService property set, which makes some of the statistics in the past 2S with the current connection in connection with the same service.
- d) host flow properties set, that host-based network traffic associated with the property, such property is designed to discover the properties of a slow scan, the approach is to obtain statistics over the past 100 connections in some of the statistical properties As in the past 100 to connect with current connection destination host with the same number of connections, and current connection with the same percentage of such service connection [9,10].

## 5 Experiment

The steps of Intrusion detection process are as follows:

- a) The interception of network data packets.
- b) To extract the characteristics of the network connection.
- c) The data pretreatment.
- d) Constructing decision information system as the input reduction algorithm.
- e) Using the feature reduction algorithm based on conditional entropy in order to optimize the decision making Information system
- f) Using the optimized sample as the input sample, training SVM sorting machine. SVM classification uses g) combination of many more than two parties Law in

order to realize multiple classification in the attacks. Constructing lots of classifiers respectively, such as DoS and Probing classifiers.

- g) Intrusion detection. After detected abnormal data, the call system would respond to the module, took appropriate response measures.

**Table 1.** Sample space

<i>The total number of samples</i>	<i>Normal number of samples</i>	<i>Abnormal number of samples</i>
2000	1780	220
5000	4420	580
10000	8960	1040

**Table 2.** Experimental results

<i>The total number of samples</i>	<i>Detection rate</i>	<i>False alarm rate</i>	<i>False negative rate</i>
2000	96.29	0.82	1.31
5000	96.60	0.69	1.14
10000	97.25	0.61	0.88

## 6 Conclusion

From the above table we can conclude that SVM algorithm has a higher detection rate, low false alarm rate and false negative rate.

In essence, Intrusion detection system is the same as a pattern recognition problem, so using the SVM method is not only feasible but also effective. Intrusion detection system is considered to be the second layer of defense behind the firewall. When working at different net levels, Firewalls using different standards to restrict the flow. Therefore, IDS will work as the same as the firewall, locates every problem in the right level of the network. The IDS system is stratified into several levels: Application Layer Intrusion Detection System (AIDS), Transport Layer Intrusion Detection System (TIDS), network layer intrusion detection system (NIDS) and link layer (LIDS), the purpose is to adapt to different levels of network attacks in order to enhance intrusion detection system performance. Each layer using its own set of feature extraction methods, and different SVM models for feature analysis, so as to get a better classification.

These different types of IDS will be distributed into different network devices. When data is transmitted on the each layer of the network, IDS will filter out malicious intrusion for each content. But this system has not considered the data sharing problems of the IDS devices in the Internet all levels , in order to better resist foreign invasion.

## References

1. Yefang, Wu, Z., Guo, L.: Intrusion of the clustering algorithm and implementation, vol. (3), 46–49. Chongqing University (2004)
2. Zhang, A. product, Xu, B.-G.: Chunking support vector clustering to improve in the intrusion detection. Micro-Computer Information 5-3(22), 46–49 (2006)

3. Lin, C., Wang, S.: Fuzzy Support Vector Machines with Automatic Membership Setting. *StudFuzz*, vol. (177), pp. 233–25 (2005)
4. Lei, H., Bin, Z., Zhou, H.: Clustering support vector machines based intrusion detection algorithm. *Radio Engineering* (39), 45–49 (2009)
5. Zaman, S., Karray, F.: TCP/IP Model and Intrusion Detection Systems. In: International Conference on Advanced Information Networking and Applications Workshops, pp. 90–96 (2009)
6. Yong-Juan, Wang, R.-C., Ren, X.Y.: SVM-based Intrusion Detection System feature weight optimization method overview. *Communication Technology*, 126–129 (June 15, 2007)
7. Bin, Z., Zhou, X.: Cluster-based fuzzy support vector machine algorithm for intrusion detection. *Journal of Information*, 175–178 (March 2009)
8. Bin, Z., Zhou, X.: Cluster-based fuzzy support vector machine algorithm for intrusion detection. *Journal of Information*, 175–178 (March 2009)
9. Scholkopf, B., Plattz, J.C.: Estimating the support of a high-dimensional distribution. *Neurral Computation* 13(7), 1443–1472 (2001)
10. Wang, L.P.: *Support vector machine:theory and application*, pp. 1–66. Springer, New York (2005)

# Application of Multi-vehicle Problem with Different Capacities

Fuxing Yang and Bowen Yu

School of Automation, Beijing University of Posts and Telecommunications,  
No.10 Xitucheng Road Haidian District, Beijing, China 100876

**Abstract.** Vehicle Routing Problem (VRP) is a typical NP-Problem which has already aroused much attention in this field. It's now an important problem in distribution industry as the delivery route influences the total cost directly and plays an import role in the whole supply chain. In this paper, we focus on the various constrains of actual scheduling problem and modeling with multi type of vehicles. We also compare the solutions with different sets of the capacity to illustrate the impact of the set of vehicles on the total cost.

**Keywords:** VRP, Multi-Vehicle Problem.

## 1 Introduction

With the development of modern logistics, Vehicle Routing Problem (VRP) is attracting more and more attentions as a typical problem of distribution system. Instances of VRP occur in various types of distribution systems, e.g., Postal Delivery and Job Shop Scheduling. Research on this problem will lead us significant effects to delivery routing optimization and help to solve other relevant problems.

VRP is a typical NP problem with multi constraints [1], there have been numerous researches on this problem. In Research on Vehicle Routing Problem Aimed at Balance, Youwang Sun proposed a new vehicle routing model based on the new concepts of logistics management that balances the workload of delivery routes [2]. In Improved Ant Colony Algorithm for Vehicle Scheduling Problems of Military Logistics Distribution, an ant colony algorithm based on the objective of minimum transportation distance is proposed to solve the multi-source-point distribution problem [3]. Hiroyuki Kawano also discussed the benefits of using GPS tracking device in Applicability of Multi-vehicle Scheduling Problem Based on GPS Tracking Records and realized dynamic scheduling [4].

## 2 Model

### 2.1 Description of VRP

The main constrains of VRP are listed below:

- a. each path has to start from the distribution center, also end at the distribution center to complete the whole delivery process.
- b. each client can be visited only once, no circle is allowed among the clients.

- c. the demand of each client has to be satisfied.
- d. the total load of each path should be less than the capacity of the vehicle assigned.

In this paper, we use the total delivery length as the evaluation function.

## 2.2 Modeling of VRP

To describe the constraints with mathematic functions, we set SITE= {0,1,...,N} represent all the spots in the system. SITE (0) is for the distribution center. SITE (1) to SITE (N) are for the clients, also set as CLIENT= {1,2,...,N}. VEHICLE= {1,2,3} stands for the vehicle owned by the distribution center.

We also set other variables as listed below.

DEM: array (CLIENT) of real, the actual demand of each client.

CAP: array (VEHICLE) of real, the capacity of each vehicle.

DIST: array (SITE, SITE) of real, the distance between every two spots.

quant: array(VEHICLE, CLIENT) of mpvar, the total amount of goods delivered by vehicle v when leaving client i.

go: array (VEHICLE, SITE, SITE) of mpvar, if go( v, i, j)=1, then vehicle v take the delivery of from spot i to spot j.

assign: array( VEHICLE, CLIENT) of mpvar, if assign(v,i)=1, then vehicle v take the delivery of client i. Assign (v,j) can also be described as below:

$$\forall v \in VEHICLE, j \in CLIENT :$$

$$assign(v, j) = \left[ \sum_{i \in SITE} go(v, i, j) + \sum_{i \in SITE} go(v, j, i) \right] / 2 \quad (2.1)$$

a. for all the vehicles that partake the delivery, the distribution path should start from and end at the distribution center, also marked as SITE(0).

$$\forall v \in VEHICLE : \sum_{j \in CLIENT} go(v, o, j) \leq 1; \quad (2.2)$$

$$\sum_{i \in CLIENT} go(v, i, 0) \leq 1; \quad (2.3)$$

b. each client can only be visited once

$$\forall i \in CLIENT :$$

$$\sum_{v \in VEHICLE} \sum_{j \in SITE} go(v, i, j) = 1, (i \neq j) \quad (2.4)$$

$$\sum_{v \in VEHICLE} \sum_{j \in SITE} go(v, j, i) = 1, (i \neq j) \quad (2.5)$$

Above two constraints can also be represented as:

$$\forall i \in CLIENT : \sum_{v \in VEHICLE} assign(v, i) = 1 \quad (2.6)$$

$$\sum_{v \in VEHICLE} \sum_{i \in CLIENT} assign(v, i) = N \quad (2.7)$$

c. to avoid circles among the clients, we should also satisfy

$$\begin{aligned} \forall v \in VEHICLE, i, j \in CLIENT (i \neq j) : \\ go(v, i, j) + go(v, j, i) \leq 1 \end{aligned} \quad (2.8)$$

d. if  $i$  is the first client of the path,  $quant(v, i) = DEM(i)$

$$\begin{aligned} \forall v \in VEHICLE, i \in CLIENT : \\ quant(i) \leq CAP(v) + [DEM(i) - CAP(v)] * go(v, 0, i) \end{aligned} \quad (2.9)$$

e. if  $j$  is the client following client  $i$ , then

$$\begin{aligned} \forall v \in VEHICLE, i, j \in CLIENT, i \neq j : \\ quant(v, j) \geq quant(v, i) + [DEM(j) + CAP(v)] * go(v, i, j) - CAP(v) \end{aligned} \quad (2.10)$$

f. the total demand of the path should be no more than the capacity of the vehicle assigned.

$$\begin{aligned} \forall v \in VEHICLE, i \in CLIENT : \\ DEM(i) * assign(v, i) \leq quant(v, i) \leq CAP(v) * assign(v, i) \end{aligned} \quad (2.11)$$

### 3 Model Output and Analysis

We assume the demand of each client and the distance between every two spots as below:

**Table 1.** Demand of client

1	2	3	4	5	6
12000	5000	4000	16000	13000	5000

**Table 2.** Distance between spots

	0	1	2	3	4	5	6
0	0	148	55	32	70	140	73
1	148	0	93	180	99	12	72
2	55	93	0	85	20	83	28
3	32	180	85	0	100	174	99
4	70	99	20	100	0	85	49
5	140	12	83	174	85	0	73
6	73	72	28	99	49	73	0

We use 3 vehicles to complete the distribution. We have to notice that not all the vehicles have to be used in the delivery.

a. if there's no limitation of the capacity, only one vehicle is needed. The shortest distance is 375, the delivery route is  $0 \rightarrow 3 \rightarrow 6 \rightarrow 1 \rightarrow 5 \rightarrow 4 \rightarrow 2 \rightarrow 0$ .

b. if the capacity is set as 40000, two vehicles are needed to complete the distribution. The shortest distance is 493. Route 1: 0→3→6→1→5→2→0; Route 2: 0→4→0.

c. if the capacity is set as 30000, two vehicles are needed to complete the distribution. The shortest distance is 504. Route 1: 0→3→2→4→0; Route 2: 0→6→1→5→0.

d. if the capacity is set as 20000, all the 3 vehicles have to partake the delivery. The shortest distance is 773. Route 1: 0→5→2→0; Route 2: 0→4→3→0; Route 3: 0→1→6→0.

We can see from above results that the capacity of the vehicle can influence the result to some extent. Besides, the evaluation function might be more completed: we have to take the steady expense into consideration, the carriage is also varies with different loads. Thud we try to use different types of vehicle to accomplish the distribution.

e. if the 3 vehicles have different capacities as 40000, 30000, 20000, then vehicle 1(CAP 40000) and vehicle 3(CAP 20000) will partake the delivery. The shortest distance is 493. Route 1: 0→2→5→1→6→3→0; Route 2: 0→4→0.

f. if capacities are set as 30000, 20000, 20000, we also have to use all the three vehicles. The shortest distance is 609. Route 1: 0→6→1→5→0; Route 2: 0→2→0; Route 3: 0→3→4→0.

We can see that plan b, c and e can provide us satisfactory results. Plan e offer us a combination of different types vehicles with different capacities, this will help to increase the flexibility of the system. Thus plan e is better than the others.

## 4 Summary

VRP is an essential problem in the vehicle dispatch system. To solve this problem, we have to consider all the relevant constrains. In this paper, we modeled and calculated this problem; we also compared the different results with different sets of the vehicle capacities.

If we take the cost of the distribution, the basic cost of the vehicles and also the flexibility of the system into consideration, the combination of various types of vehicles will serve better than a single type of vehicle.

## References

1. Timucin Ozdemir, H., Mohan, C.K.: Evolving Schedule Graphs for the Vehicle Routing Problem with Time Windows. In: Evolutionary Computation, pp. 888–895 (2000)
2. Sun, Y., Song, H.: Research on Vehicle Routing Problem Aimed at Balance. Logistics Sci.-Tech. 6, 22–25 (2010)
3. Gong, Y., Huang, R.: Improved Ant Colony Algorithm for Vehicle Scheduling Problems of Military Logistics Distribution. In: 2010 International Conference on Logistics Systems and Intelligent Management, pp. 669–673 (2010)
4. Kawano, H.: Applicability of Multi-vehicle Scheduling Problem Based on GPS Tracking Records. In: 2010 International Conference on Geoinformatics, pp. 1–4 (2010)

# **Study on Risk Control of Network Transactions Based on Customer Perspectives**

Mengting Sun

International Institute, Beijing University of Posts and Telecommunications, Beijing, China  
linghuchong\_2008@126.com

**Abstract.** This paper proposes a risk evaluation model to solve the problems of risk control in network transactions. Possible risks in network transactions during the payment activities are recognized and analyzed, and different risk evaluation index systems and risk evaluation models according to different corresponding risk are discussed. Then, some corresponding risk control and management measures are put forward. During the process of research and analysis, the corresponding control methods of various risks based on customer perspective are put forward.

**Keywords:** network transactions, risk control, customer perspective.

## **1 Introduction**

Along with the development of information technology, electronic and networked era comes up. E-commerce brings tremendous change to our life and makes our life convenient [1]. But it also brings us some new problems. As buyers and sellers don't meet each other in trading process, and they are lack of mutual understanding, so there could be some fraud behaviors during payment non-delivery or delivery.

Hence, the network payment platform emerged, and it solved the security problem of online transactions, promoting the development of electronic commerce. But along with the continuous development of network pay industry, some money laundering events appeared which reflects certain risks. The network transactions have emerged in recent years, so it has not been extensively researched. This is what this study its meaning.

## **2 Classifications in Network Transactions**

### **2.1 Safety Problems**

- Electronic pickpocket

Some people are called "electronic pickpocket", who imitate special steal others network address. This kind of theft has rapidly rising trends in recent years. As the Internet services provide banks and users shared resources, but also provide chances to steal banking. Some thieves steal bank or business secret, or browse enterprise core secrets.

- Internet fraud

Internet fraud has become the second most common network risk. Some outlaws sent through email or other kinds of attractive free material, etc. When a user use these emails or materials into the bank's web site, the imbedded modified software will automatically put on the user account and transfer money into lawlessness molecular account.

- Computer hackers

Generally, illegal invaders of the computer system are called "hackers". This concept is proposed by the Massachusetts institute of technology's scholar. At present, there are many countries have manufacturing electronic bomb ability, which make the state's financial security of potential risks more great.

- Computer viruses

Computer virus has formed great threats on the e-commerce. The known computer virus in the world has reached to hundreds of 18,000 and also many are unknown. In addition, with the popularization of the Internet electronic mail, the trade electronic document has become the main channel of computer virus spread.

- Information pollution

Just as the industrial pollution during industrial revolution, the information age also have information pollution and information excess. Lots of unordered information is not resources but disaster. As Internet subscribers and network traffic has increased dramatically, many new problems appears, such as large advertisements online "junk"[2].

- Operational issues

Operational problems mean potential losses because of the big defects of system reliability and stability. This kind of problems may come from the network bank security system and its product design flaws. These risks mainly conclude the network bank risk management system, the network bank information communication with clients and chest e-money recognition, etc.

## **2.2 Moral Hazards Caused by Information Asymmetry**

Due to network transactions on the Internet is not transparent, it is difficult to identify the level of risk and make objective customer in disadvantaged choice position. Online customers could use their hidden information and make concealed action favorable network bank. In addition, in virtual financial markets, each bank online customer don't understand what is the quality of service provided by high and low, most customers will follow their provide services to the network bank to determine the average quality expected the purchase price. As a result, high quality network bank may actually be crowded out the network market by low quality of the network bank [3].

## **2.3 Lack of Relevant Laws and Regulations**

The network law risks violation comes from relevant legal provisions, rules and regulations. The network transactions are still in the initial stage, the government does not have necessary, complete laws and regulations of financial legislation framework to adapt. Although countries have the relevant laws and regulations, but network is across borders. In the network multinational trading business, unavoidable legal problems between countries of the conflict will be produced.

### 3 Choice of Risk Evaluation Methods in Network Transactions

#### 3.1 Credit Evaluation Method

There are a lot of evaluation methods for credit risk, such as traditional credit evaluation method based on discrimination models, Z score models. In addition, such as KMP, genetic programming, bayesian network, neural network evaluation model, are all can be used in the credit risk assessment. But relative to the different methods, they have their more applicable fields [4].

A risk measurement tool originally generated based on certain types of risks. But after a period of development and improvement in other areas, it developed more effectively. For example, as risk management method, VAR initially is mainly used to market risk measurement. But after development and constantly improvement, it also can be extended to quantitative research on credit risks.

#### 3.2 Cash Flow Evaluation Method

VaR is a kind of financial up risk assessment and econometric model. VaR (Value at Risk Value or associated) is called risk value, meaning for at-risk value. It means the utmost expectation loss in normal market conditions, when risk assets or combination in a given confidence interval and holding horizon.

The specific calculation formula is:

$$\text{prob}(R \leq \text{VaR}) = 1 - \alpha$$

R means expected loss, and VaR is risk value.  $1 - \alpha$  is the given credibility, which means preferences of risk degree, generally take 90 to 99.996 value. VaR model summarizes the risk as simple currency values, which represents the potential future risk said loss rate, rather than merely a probability numerical. And it also includes the risk of investment losses and the probability of occurrence in loss rate.

#### 3.3 The Fuzzy Comprehensive Evaluation Method

This method is introduced based on the nonlinear characteristics and evaluation process of fuzzy mathematics and the quantitative evaluation of the results can be get be virtue of the fuzzy algorithm for nonlinear evaluation. It is widely used in various fields, and it is a kind of very effective multi-factor decision-making method. This evaluation model help people keep thinking process consistency for ministering in fuzzy comprehensive evaluation of various factors in determining weights. The specific calculation steps are as follows.

##### 1) Establish factor sets

Set of factors by n constitute factors set. U usually means  $U = \{u_1, u_2, \dots, u_n\}$ , which represents type of various factors, and the factors often have varying levels of ambiguity.

##### 2) Establish weight sets

The  $u_i$  ( $i = 1, 2, \dots, n$ ) is give the corresponding weight  $a_i$  ( $I = 1, \dots, n$ )and they meet normalization conditions.

### 3) Establish evaluation sets

Set of evaluation consists of results judged by m kind assessment. This set is usually composed of V, that is  $V = \{v_1, v_2, \dots, v_n\}$ , which represents each possible evaluation results.

### 4) Evaluation

Fuzzy sets in fact can be considered as a fuzzy relation, which represent factor sets U and evaluation sets V between fuzzy evaluation sets. Therefore it can be expressed as,

$$R_i = \frac{r_{i1}}{(u_i \cdot v_1)} + \frac{r_{i2}}{(u_i \cdot v_2)} + \dots + \frac{r_{im}}{(u_i \cdot v_m)}$$

5) Processing evaluation results and the final evaluation results will be obtained.

### 3.4 Numerical Example

Consider the network risks consist of five factors. U means  $U = \{u_1, u_2, u_3, u_4, u_5\}$ , and the value can be got by expert system, that is  $U = \{0.16, 0.20, 0.20, 0.32, 0.23, 0.09\}$ .

By virtue of the fuzzy evaluation model, we can get the final comprehensive appraisal matrix:

$$R = \begin{bmatrix} 0.4 & 0.16 & 0.18 & 0.19 & 0.05 \\ 0.22 & 0.23 & 0.19 & 0.29 & 0.063 \\ 0.34 & 0.34 & 0.12 & 0.15 & 0.045 \\ 0.44 & 0.23 & 0.14 & 0.17 & 0.03 \\ 0.22 & 0.13 & 0.34 & 0.18 & 0.12 \end{bmatrix},$$

$$A = U \circ R = \{0.33, 0.23, 0.17, 0.21, 0.05\}.$$

According to this result, we can conclude that  $u_5$  means the minimal risk.

## 4 Measures for Risk Control on Network Transactions

### 4.1 Internal Control Measures of Risks

#### 1) Choosing the right payment business

For consumers and businesses are concerned, they should choose to have credibility of the third-party providers. These providers have reputable service, and they have effective risk prevention system conditions, etc.

#### 2) Strengthening the secure payment awareness

The customers should strengthen secure payment consciousness and develop security habits. Do not install the unexplained software, and don't open the unexplained E-mail. Withdraw payment from third-party platform bank account, and do our best to avoid possible risks.

## 4.2 External Control Measures of Risks

### 1) Strengthen supervision

It is necessary to formulate the third-party market access rules, such as setting a minimum capital restrictions, strengthening internal control mechanism and risk management, strengthening the safety technology, establishing insurance and deposit system, etc.

### 2) Standardized construction

The network transactions contractors should establish a set of industry service standards through consultation to clear their obligations and responsibilities. And they should take the good faith construction as the key point to protect the interests of consumers. Uniform industry standards and norms should be established to increase the convenience for consumers [5].

### 3) Develop social network environment

Social network environment foster mainly includes cultivation of honesty and paid service consciousness. Internet users can support "Internet trust plan" and "Internet trust plan" to cultivate the Internet playing a positive role.

## 5 Conclusions

This paper studies the corresponding evaluation of the risks in network transactions and puts forward related risk prevention methods. In the future of subsequent research, more detailed analysis should be researched and more concrete, more effective risk control schemes should be given.

## References

1. Aohui: Credit guarantee project risks based on multi-grade fuzzy comprehensive evaluation. Journal of Wuhan University of Technology (10), 121–123 (2006)
2. Penza, P., Bansal, V.K.: Measuring Market Risk with Value at Risk. John wiley & Sons, Ltd. (2001)
3. Basak, S., Shapiro, A.: Value-at-Risk-Based Risk Management: Optimal Policies and Asset Prices. The Review of Financial Studies 14 (2001)
4. Christoffersen, P., Jinyong, H.: Testing and comparing Value-at-Risk measures. Journal of Empirical Finance (8), 325–342 (2007)
5. Schnoider, G.P., Perry, J.T.: Electronic Commerce. China Machine Press, Beijing (2008)

# FPGA-Based Design and Implementation of Video Wall Display

Yang Li<sup>1</sup> and Xuesen Cai<sup>2</sup>

<sup>1</sup> Jilin Architectural and Civil Engineering Institute, 130118, Changchun, China  
liyang9750@126.com

<sup>2</sup> College of Computer Science and Technology, Changchun Normal University, 130026,  
Chang Chun, China  
oldcai@126.com

**Abstract.** With the rapid development of science and technology, industry, agriculture, transportation, etc the demand for image output equipment has increased, therefore a large number of high cost big screen displays are required. In order to reduce costs, relative to the existing electronic product cost, the joining together, display technology division, the design that can satisfy the demand of the monitor or a combination of the alternative splicing wall. This paper expounds the design process, monitor hardware structure and the logical structure, eliminate a closer look at the serrated feeling, improve when enlarge display screen.

**Keywords:** FPGA, Built-up Wall Display, Small Point, Feeling of Serrated.

## 1 Introduction

The monitor is the part of the monitoring system, used to display the image. Through observation monitor real-time monitoring personnel to judge the picture of the security situation monitoring targets. Monitor in the image of clarity, color reduction degree, working stability, has high performance index, monitoring and control system is an important part.

According to the principle that, monitor can be divided into CRT (cathode ray tube), plasma and LCD monitor, etc. CRT monitors color reduction degree is high, the large size, small size display; Plasma monitor the brightness, contrast, high, show size, but its weakness is static pictures show to "burn point"; LCD monitor high-definition, display size and low power consumption. At present the CRT monitors are gradually be replaced LCD monitor.

## 2 Technology Research Situation at Present

The monitor industry generally uses less than 25 inch monitors, the most common being 17 to 22 inches. This size of display device utilises a low resolution, generally 1024 \* 768 or 1440 \* 900. The main benefits of smaller size monitors are the clarity

of the picture and the relative low cost, The main disadvantage being that the display is too small when viewed from a distance, with details being lost, due to the low resolution, show more screen resolution will only display the reduced signal. And the source image shows the reduced, may lose the details of the picture, and monitor the operation of the object of monitoring impact assessment of the state.

The development in photoelectric technology, has lead to digital cameras being produced that are capable of taking pictures of more than 15 million pixels, but high end display devices are currently limited to about 2 million pixels (1920 \* 1080), if need to complete the picture show, only will pictures reduced proportion, need to see the details, and partial enlargement. In this way, will complete display enables small details of the more difficult to observe, local and not to show the whole evaluation.

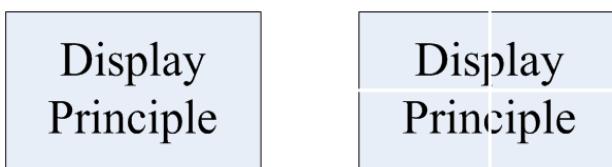
The paper presents the design of a high resolution monitors, can at the same time satisfy the large size and small point from which the requirements.

### 3 System Overall Designs

This paper presents the design of high resolution monitors, including 1080 P60Hz capable of receiving all kinds of common video signals and displaying them on an LCD panel, in proportion to that of 8 million pixels draw pictures, screen size for 82 inches, has a very high resolution and larger visual size.

#### 3.1 Show Unit Timing Requirements

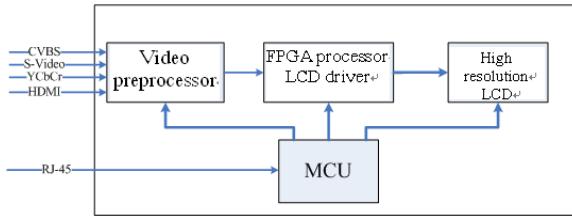
Display module USES the south Korean a company the production of 82 inch LCD screen resolution, 3840 \* 2160, effective pixel about 8 million. Liquid crystal screen interface for two Dual-link DVI interfaces, through each DVI interfaces receiving data, internal drive circuit driver and a half of the screen display area, namely 1920 \* 2160, shows driving way shown in figure 1.



**Fig. 1.** Effect of mosaic display

#### 3.2 System Hardware Diagram

At the front end the video signal processing system, followed by the FPGA storage drive circuit and the liquid crystal screen, all controlled by the MCU, a total of 4 main components.



**Fig. 2.** System hardware diagram

Because video monitoring equipment and applications of different, monitor equipment will probably send out NTSC, PAL, HD, Full-such as HD video signal of different formats, must be compatible with the common monitor all video format. This paper front video processing circuit should be able to handle various simulation and digital video signals, including the 480 I, 576 I, 720 P, 1080 I and 1080 P video format of format, will do to deal with, such as, amplification interlaces change into the 1920 x1080x60Hz video signal to the FPGA. So when the FPGA design only to a form of signals do processing, reduced design of difficulty. The FPGA received 1920 x1080x60Hz signal or MCU of image data sent, and stored in memory, and according to the LCD screen calendar will display the data driven from DDR2 memory to read LCD. Because of the high resolution LCD screen, which requires the FPGA has strong data processing, the cache ability.

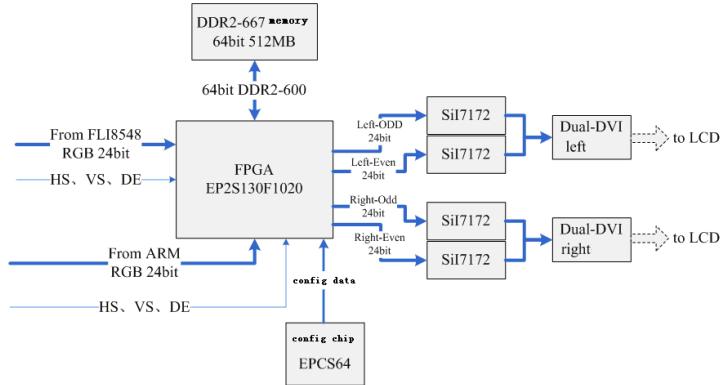
## 4 System Hardware Design

### 4.1 Front Video Signal Processing Circuit

FLI8548 will X86 CPU, RGB/YPbPr A/D converter, 3 video decoder, where the HDMI decoder and take the motion compensation high performance in A single integrated zoom controller chip can eliminate the standard interlaces the video signal through the display equipment manufacture progre ive-scan when deformation of image edge. Because FLI8548 no audio processing functions, so vast a SGTV5810 audio processing chip, to the input channel simulation audio signal and FLI8548 decoding out HDMI digital audio signal processing. SGTV5810 SigmaTel company is the production of TV audio processing special chip, it supports simulation and digital audio signal input and output. Input interface including 9 groups stereo simulation, a group of audio input interface I2S digital input and a synchronous asynchronous I2S digital input interface group. Output interface including 5.1 channel simulation signal output and SPDIF digital output interface and asynchronous I2S output interface. Through the I2C interface to control its working mode.

### 4.2 FPGA and Image Storage

The FPGA hardware structure as shown in figure 3 shows:



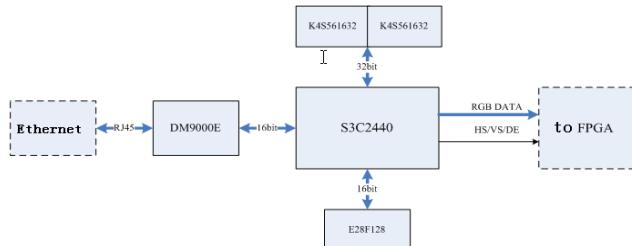
**Fig. 3.** FPGA and image storage, liquid crystal drive circuit

System of FPGA realizing function is more complex, each part of data transmission frequencies are higher, so the choice of system based on FPGA Altera company Stratix II series of EP2S130F1020. And use EP2S130 Altera company with the DDR2 \_Controler IP Core, can achieve the highest for DDR2 667 stimulation rate. In this paper the actual operation rate of DDR2 600. Because EP2S130 an SRAM structure for the FPGA, not after power off, it saves the program a EPCS64 plugins. EPCS64 is based on the structure of the FLASH Altera FPGA program configuration chip, after power up can offer EP2S130 program loading sequence, automatic configuration EP2S130.

#### 4.3 Main Control MCU Circuitry

The MCU controls various parts of the system, including receiving the transmission, converting this to pictures and decoding the data for the FPGA. See figure 4 below.

The main control MCU samsung's S3C2440 ARM9 processor. This type of processor as ARM920T kernel, main frequency for 400 MHz. Outside enlarge two pieces of K4S561632 32 bit, composed of 64 MB SDRAM memory, the use of a E28F128 FlashROM program memory, as after power up by Flash read the program to load the SDRAM and operation. Because S3C2440 interior does not have the Ethernet interface, so vast a DM9000E. DM9000E is a 10 M / 100 M adaptive Ethernet interface chip, with compatible with ISA bus interface, which can directly and S3C2440 articulated. S3C2440 have internal memory management unit (MMU), the paper 2.6.22 Linux operating system running on the S3C2440.

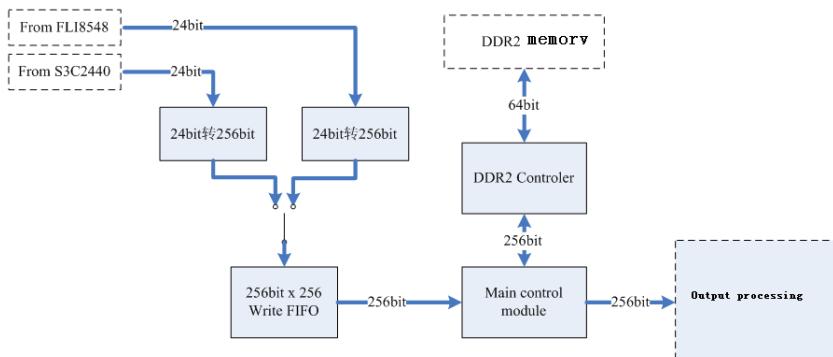


**Fig. 4.** MCU control circuit

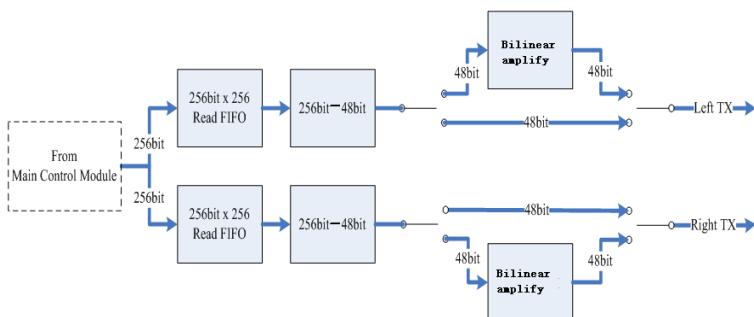
## 5 The FPGA Logic Function Design

### 5.1 The Whole Logic Structure

The FPGA main function is to receive video processing circuit or ARM of the data sent to DDR2 memory, and at the same time, to read the data stored in DDR2 and according to the timing driving LCD screen output for two road 1920 x2160x60Hz timing. As the clock frequency different input and output part, DDR2 local BUS working frequency for 150 MHz, 1080 P60Hz data input frequency of 148.5 MHz, ARM to the frequency of FPGA data about 20 MHz, output to LCD screen the frequency of data for 138 MHz. So in this article, the three FIFO, realized respectively the input continuous data writing DDR2, will be stored or so two and a half screen display data to a 127 MHz frequency sent out to LCD screen. Integral logical architecture as shown in figure 5 below.



**Fig. 5.** Overall logic structure diagram



**Fig. 6.** Output processing structure

Input part through the data selector choose receive FLI8548 or ARM delivery of the data. Signal through "24 bit to turn 256 bit" modules, each receiving 10 points of data a written after RGB Write\_FIFO. Main\_control\_module as the core control module, responsible for reading input data, to read data to the DDR2, output FIFO. Its function essence is a circulating state machine.

Output processing has two functions, if it is the ARM of the sent high resolution image is directly flooding screen display; If it is FLI8548 sent 1080 P60Hz image, is still need to be amplified module, and whole hd image amplifier to four times the resolution of the high definition, then send them to the display on the LCD panel.

## 5.2 Full hd Video Magnification Shows

LCD display resolution for 3840 x2160, and full hd video input signal for 1920 x1080 resolution, so each frame of the hd image in horizontal and vertical direction were magnified for the original 2 times, can full screen. Considering the resolution of the input signal is higher, the color of the two adjacent data is not instantaneous mutations, so in this using simple bilinear interpolation algorithm, in horizontal and vertical direction using two linear interpolation amplification. Specific as follows: A horizontal display data: D11, D21, D31..... Enlarge image data are for: D11, (D11 + D21) / 2, D21, (D21 + D31) / 2, D31..... A vertical display data: D11, D13....., D12 Enlarge image data are for: D11, (D11 + D12) / 2, D12, (+ D13 D12) / 2, D13.....

## 6 Last Word

This paper used the FPGA and high-speed storage device design a new high resolution monitors, its ultra high display resolution in the display high pixel image content at the same time, not lost details, can be widely used in transportation scheduling, aviation shooting, railway and monitoring that higher requirements for details of the field.

**Acknowledgments.** The authors wish to thank the organizing committee for providing this chance to communicate each other. And will thank doctor Fanhua Yu for his technical support.

## References

1. Liu, J., Niu, Y.X.: Design and application of video signal generator based on FPGA. Chinese Journal of Scientific Instrument 29(3), 654–657 (2008)
2. Jiang, Y., Gu, T.X.: Quick frequency estimation based on MUSIC algorithm. Chinese Journal of Scientific Instrument 27(11), 1526–1528 (2006)
3. Hu, Y.F., Chen, H., et al.: Design and implementation of model predictive controller based on FPGA/SOPC. Chinese Journal of Scientific Instrument 31(6) (2010)
4. Yang, Z.H., Zhou, P., et al.: Improvement and implementation of the algorithm design of elliptic curve dot product based on FPGA. Chinese Journal of Scientific Instrument 30(7), 1546–1551 (2009)
5. Wang, Y., Tan, H.: Design and implementation of data acquisition and transmission system for industrial CT. Chinese Journal of Scientific Instrument 29(4), 722–727 (2009)
6. Chong, T.S., Au, O.C., Chau, W.S.: A content adaptive de-interlacing algorithms. In: IEEE International Symposium on Circuits and Systems, vol. 5, pp. 4923–4926 (2005)
7. Li, M., Nguyen, T.: A de-interlacing algorithm using markov random field model. IEEE Transaction on Image Processing 16(11), 2633–2648 (2007)

# Research on B-tree in Embedded Database SQLite

Di Nan

Computer Science and Technology, Harbin Engineering University  
Harbin, China  
lijingmei@hrbeu.edu.cn

**Abstract.** The methods that format the data in database and organize the structure of B-tree-page in open-source embedded database SQLite is described, and the database which is established by SQLite is analyzed. Further, the code of achieving this part of SQLite is analyzed in depth. Finally, the code to operate the red-black tree is designed, which uses the interface function provided by SQLite, then realizes the application of B-tree in the database.

**Keywords:** Database, B-tree-page, SQLite.

## 1 Introduction

Embedded database SQLite runs in the process of application program directly, so SQLite can work in the zero-configuration mode and occupy little resources. SQLite is realized by C language, which provides full independence and openness even does not depend on other resource out of the application program. SQLite is introduced as one of PHP V4.3 options and built on PHP V5 [1]. SQLite supports SQL92 standard and can be ported to all major embedded operating systems, even supports a lot of major high-level languages (such as C and Java, etc.). In addition, SQLite is also good robust, and can handle 10,000 CTR on Web site everyday[2]. The size of database file that is created by SQLite is up to 2 TB, and each database can store in the single file system entirely. These files can be transported between the computers even in the different byte order. These data are stored on disk in the form of B-tree, and SQLite gets the authority to access its database from file system. If SQLite is used to manage the database, the first step is to analyze the method how SQLite realizes B-tree.

## 2 The Structure of B-tree

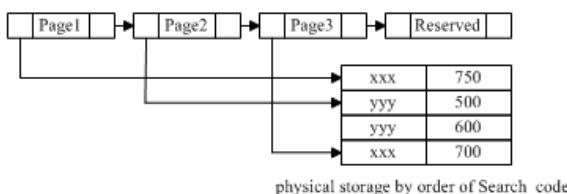
B-tree was invented by R. Bayer and E. McCreight in 1970, and its structure is different from the multi-index formed by order index[4]. All the nodes in B-tree have the same structure as shown in Figure 1, which searches the node with the number of  $n-1$  (such as  $K_1, K_2, \dots, K_{n-1}$ ) and the pointer with the number of  $n$  (such as  $P_1, P_2, \dots, P_n$ ), every node in the search codes is stored in order. Pointer  $P_i$  ( $i = 1, 2, \dots, n-1$ ) points to the pointer barrels that have search code  $K^i$ , and any pointer in the pointer barrels points to file with the search code  $K_i$ .



**Fig. 1.** The structure of node in B-tree

## 2.1 Leaf Node in B-tree

The search code of each leaf node is in the number of 1 to n-1, and it is different from each other. If index of B-tree is fully indexed, any search code in database files must be in a leaf node only once. In addition, the leaf nodes are sorted to linear sequence by the value of search code, so SQLite can use the pointer Pn of each leaf node to connect the leaf nodes orderly, as is shown in Figure 2, which makes B-tree efficient to operate the database file in linear.



**Fig. 2.** Structure of leaf nodes in B-tree with n = 3

## 2.2 Non Leaf Node in B-tree

B-tree uses the non leaf nodes to form the multi-index of the leaf nodes, and the structure of the non leaf node is the same as the leaf node. No leaf node also contains search code to form the structure of storage unit, However, all the pointers of the non leaf nodes point to the nodes in B-tree. A non leaf node contains pointers with the number of m ( $1 \leq m \leq n$ ), if  $m < n$ , from  $P_m$  all the free space is reserved. For example, a non leaf node have pointers with the number of m, the pointer  $P_i$  ( $1 < i < m$ ) points to a sub-tree, In the nodes of this sub-tree all the search code that are less than  $K_i$  are equal to or greater than  $K_{i-1}$ . Pointer  $P_m$  points to the part of sub-tree that contains the search code greater than or equal  $K_{m-1}$ , and the pointer  $P_1$  points to the part of sub-tree that contains the search code less than  $K_1$ .

# 3 The Structure of Database File

If there is a lot of data in the database that is organized by B-tree, it is much faster to achieve the operation such as searching, removing and adding. But the structure of large database is very complex, so in this paper a simple database that can be analyzed easily is established, besides this database file only has only one table.

### 3.1 The Establishment of the Database File

Firstly, a simple database is created by SQLite:

Create table sample (one varchar (10), two varchar (10));

Insert into sample values ('1 st ',' xxx '); Insert into sample values ('2 nd ',' yyy ');

Insert into sample values ('3 th ',' zzz ');

If the following order is implemented in SQLite, the data can be queried from the table.

Select \* from sample; And the table is shown like that.

1st | xxx; 2nd | yyy; 3th | zzz

Although there is only one table in the database, two B-tree pages are in the database file, SQLite defines that the size of each B-tree page is 1 KB, so the size of this database is 2 KB. The B-tree page is the same as the classic structure of B-tree node , because in SQLite B-tree is defined and realized as the page.

### 3.2 The First Page of the Tree

The machine code of the first page in B-tree is shown in Figure 3:

```

00000000h: 53 51 4C 69 74 65 20 66 6F 72 6D 61 74 20 33 00 : SQLite format 3.
00000010h: 04 00 01 01 00 40 20 00 00 04 00 00 00 00 00 ; .....8 .....
00000020h: 00 00 00 00 00 00 00 00 00 00 00 01 00 00 00 01 ; .....
00000030h: 00 00 00 00 00 00 00 00 00 00 00 00 01 00 00 00 ; .....
00000040h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ; .....
00000050h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ; .....
00000060h: 00 00 00 00 00 00 00 01 03 B2 00 03 B2 00 00 ; .....7.?
00000070h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ; .....
.....
000003a0h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ; .....
000003b0h: 00 00 4C 01 06 17 19 01 75 74 61 62 6C 65 73 ; ..L.....rtables
000003c0h: 61 60 70 6C 65 73 61 60 70 6C 65 02 43 52 45 41 ; amplerample.CREX
000003d0h: 54 45 20 54 41 42 4C 45 20 73 61 6D 70 6C 65 28 ; TE TABLE sample(
000003e0h: 6F 6E 65 20 76 61 72 63 68 61 72 28 31 30 29 2C ; one varchar(10),
000003f0h: 74 77 6F 20 76 61 72 63 68 61 72 28 31 30 29 29 ; two varchar(10))

```

**Fig. 3.** The data of the first page

The first page of B-tree is from 0000h to 0400h, the information of the table SQLite Master which is the system table of SQLite database is stored in size of 100Byte after the address 0000h. These data only exists in the first page of B-tree, and other pages are not formed by this structure, so the first page is always defined as the head. These binary uses the method of big-endian, the specific definitions are shown in tables 1 [6].

Behind that storage area that occupied by SQLite Master is: head-page structure of B-tree, pointer structure of B-tree, unused space and actual data load of B-tree. These structures are same in each page.[5].

The classical structure is for the purpose of aiming the clear principles of B-tree, os there are some differences from the classical structure of B-tree for the purpose of application easily, In the general page of B-tree structure is that: pointer, data, pointer, data, pointer, ..., data, pointer. but SQLite organized the page like this: pointer, pointer, ..., pointer, data, data, ... data.

**Table 1.** Structure of file's head

1. offset	1. size	1. definition
2. 0	2. 16	2. SQLite version 3
3. 16	3. 2	3. The size of each page is 1024 KB
4. 18	4. 2	4. version
5. 20	5. 1	5. The data is stored from the last byte
6. 21	6. 1	6. The maximum number of patch when the page is over load
7. 22	7. 1	7. The minimum number of the patch
8. 23	8. 1	8. The minimum number of the patch in the leaf node
9. 24	9. 4	9. Counter of amending file, facilitate to parallel visit
10.28	10.4	10. retain word
11.32	11.4	11. The first free table
12.36	12.4	12. The number of free table
13.40	13.60	13. No use

The structure of head in the first page is defined from Address 0064h, its specific definition is shown in Table 2. [6].

Pointer structure of B-tree starts from 006ch, which only has one pointer that points to the offset 03b2h in this paper. This area stores actual data load of the database. Actual data load of SQLite Master table is from 03b2h to the end of page, these data is also organized in rules. 4c in the 03b2h means that the size of records is 76Byte, and the number of index in 03b3h is 01. Following that is data load of PAYLOAD, the size of entire recorded is 76Byte as expressed in front.

**Table 2.** Structure of page's head

1. offset	1. size	1. definition
2. 0	2. 1	2. 1:intkey;2:zerodata;4:leafdata;8:leaf
3. 1	3. 2	3. The offset of first free block is 0
4. 3	4. 2	4. Create only one table: sample, thus the record is 1
5. 5	5. 2	5. The first address of load area is 03b2h
6. 7	6. 1	6. The number of patch. Data is little, so set 0

### 3.3 The Second Page of the Tree

Following the first page is the second page of B-tree, which is shown in Figure 4.

Because it is not the first one, there is no 100 Bytes header like first page. Thus the first data in the second page is the structure of page-head. There are two pointers 03F3 and 03E5, and the other is the same as the first page. Managing the entire system of database is to manage every page like that.

```

00000400h: 00 00 00 00 03 03 DF 00 03 F5 03 EA 03 DF 00 00 2 .....??
00000410h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 2 .....??
.....
000007d0h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 2 .....??
000007e0h: 03 03 13 13 33 74 68 7A 7A 7A 09 02 03 13 13 32 2 ....3thess...?
000007f0h: 6E 64 79 79 79 09 01 03 13 13 31 73 74 7B 7B 7B : mdyyy....latxxx

```

**Fig. 4.** The data of the second page

## 4 Achieving B-tree in SQLite

In SQLite, btree.c and btree.h is used to manage the B-tree.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

### 4.1 btree.c

In btree.c, three structures are defined.

```
struct MemPage; struct Btree;struct BtCursor;
```

Basic functions such as searching, adding and removing the node is also defined in btree.c.

Each table of the database is stored in the memory as the structure of page. In MemPage, adisk [] is used to store the origin data read from the memory, but other supporting information which is only efficiency to the standard pages is still stored in the memory. The overflow pages and the pages on the free list are not operated in this way. These information is the access to apCell [] that points to the basic unit of adisk []. The space of apCell [] is large enough for two units, and user's program can visit the unit stored in the array efficiently. In general all the apCell [] points to aDisk [], but in the operation of inserting some apCell [] points to the space out of adisk [] temporarily.

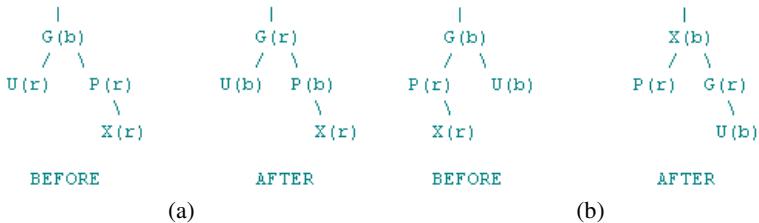
### 4.2 Application of B-tree

Red-black tree is a special B-tree, each node of red-black tree is infected by red or black in rules: all of the codes have two colors, red or black, the leaf node is black, if a node's color is red, then the son node must be black, there are the same black nodes from any specified node to its offspring nodes. Edit the user program rbtree.c, call the interface function in btree.h to realize red – black tree.[4].

Node pX can be insert into B-tree pTree by function sqliteRbtreeInsert (), this process can be archived by calling the function in btree.h. But the node inserted into the red-black tree maybe have the father node in red, which does not comply with the attributes of red-black tree. Therefore, it is necessary to rotate the B-tree and change the color of some nodes, so that it would make the B-tree return balance. If the parent node of pX exists and its brothers are red, transforming as following,

```
pGrandparent-> isBlack = 0; pUncle-> isBlack = 1;
pX-> pParent-> isBlack = 1; pX = pGrandparent;
```

This Specific situations is shown in Figure 5. (a)

**Fig. 5.** Situation of transformation

If pX is the right child, firstly it should be transformed to the left child.

$pX = pX \rightarrow pParent; leftRotate(pTree, pX);$

Then, transforming as following, which ensures the balance of the tree,

$assert(pGrandparent == pX \rightarrow pParent \rightarrow pParent); pParent \rightarrow isBlack = 0;$

$pX \rightarrow pParent \rightarrow isBlack = 1; rightRotate(pTree, pGrandparent);$

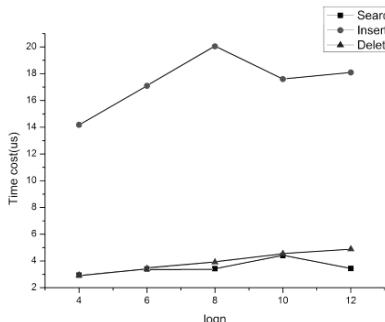
The specific situation is shown in Figure 5(b).

### 4.3 Application of B-tree

In order to evaluate the red-black tree, we run the test program on the system of Intel(R) Core(TM) 2 Duo CPU. In experiments, we achieve various operations of red-black tree such as: SEARCH, INSERT, PREDECESSOR, SUCCESSOR, MINIMUM, MAXIMUM, DELETE and so on. The testing result is shown in table 3, the time unit is  $\mu s$ .

**Table 3.** Structure of page's head

1. nodes	1. search	1. insert	1. delete
2. $2^4$	2. 2.83	2. 14.15	2. 2.83
3. $2^6$	3. 3.38	3. 17.05	3. 3.39
4. $2^8$	4. 3.41	4. 20.02	4. 3.89
5. $2^{10}$	5. 4.38	5. 17.55	5. 4.52
6. $2^{12}$	6. 3.42	6. 18.05	6. 4.85

**Fig. 6.** Data comparison chart of red-black algorithm

According to the data in table 3, we draw a graph (Figure 6) which show the three ways. There are more obvious effects of n when the tree is small, in the three measurement curves, insert is more match of the growth rate of lgn. When the n is small, the execution time of algorithm impact by the configure of the system.

## 5 Conclusion

Research and develop on B-tree-page of open-source embedded database SQLite would ensures that the user database can be operated more efficiently, and adapt to the different requirements of user's database in the embedded system.

## References

1. Zhang, G.: PHP5 programming. Electronics Industry Publishing House, Beijing (2007)
2. Chen, Q., Zhao, Z.: The Ruby Way. Posts and Telecommunications News Press, Beijing (2007)
3. Jung, K., Brown, A.: Beginning Lua Programming. Wroclaw Press (2007)
4. Feng, Y.: Data structure and algorithm analysis: C language description. Machinery Industry Press, Beijing (2004)
5. Wang, S.: Research and application of ARM-based embedded database. Micro-Computer Information 23(10-2), 62–64 (2007)
6. Bai, T., Chen, Z.: Construction of data acquisition and dissemination in Embedded system. Computer Engineering 33(19), 270–272 (2007)

# The Method of Parallel Design Based on Simulation

Jingmei Li, Qi Zhang, and Nan Di

Computer Science and Technology, Harbin Engineering University,

Harbin, China

{lijingmei,s310060097,lijingmei}@hrbeu.edu.cn

**Abstract.** Beginning with the analysis of the limitation which the sequence design of hardware prior and the parallel design of hardware and software both have, do some researches on the method of parallel design based on simulation which ensures the balance of hardware and software, rapid prototyping design and avoids the risk by simulating on testing platform in design process of the embedded system,. Through the analysis and design of a framework of simulation platform based on ARM, illustrate the application of simulation platform in this method.

**Keywords:** Algorithm of hardware and software prototyping, System simulation, Structure of embedded system.

## 1 Introduction

Embedded systems were more and more popular, but "embedded systems" was not the latest concept, which was born with the generation of computer. Then the development of computer was in parallel stages both on the development of common computer and the embedded system which was used in the target system as intelligent controller [1]. From the success of the 8 bit MCU to the application of 32 bit ARM, the design method of embedded system was different fundamentally, embedded system developers had to change the method and meet the requirements such as low cost, high performance and so on. Since the progress of electronic science and technology, and electronic product's life cycle was shorter and shorter in market, it was very important to minimize time of development. But embedded system design still used the method that divided the system into basic hardware and specific software:

1) To the hardware design of special application: the ASIC was used in this part before. Because of the characters of electronic products and the restrictions of ASIC, components of IP were accompanied, which was similar with the software and possible to repeat using hardware components. IP components allowed developing systems by FPGA to replace ASIC in many areas [2].

2) To the software both in the specific application and the OS which communicated with the embedded microprocessor: the high performance of hardware simplified the design of systems now, which allowed RTOS running in the embedded system fundamentally, and made embedded systems enhancing the overall performance and the complexity of function rapidly, but the real-time system required rapid response time, so the task switching mechanism of embedded OS in the low level software played an important role in the performance of the system.

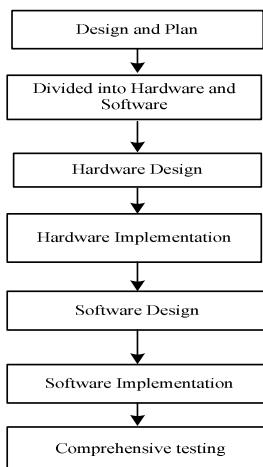
To sum up, because hardware and software design method had been more mature, in order to improve the speed and quality of embedded system design, which needed more balance and low risk approaches in separation and synthesis of hardware and software.

## 2 Traditional Process of Embedded System Design

With the advancement of electronic technology, the design of embedded system had experienced two phases of sequential design and parallel design.

### 2.1 The Hardware Prior Sequential Design

Fig.1 showed that in sequential design the separation steps of hardware and software was easy and saved time, but it was based entirely on the designers' intuition and could not be guaranteed at the beginning of the design, developers could only verify the accuracy of hardware and software at the final step. If there were errors which could not be avoided in the process of connecting the two parts, the design must be restarted. In addition, because of the restrictions of developer's knowledge and experience, for reducing errors, they selected the components that were used frequently, which reduced the freedom of choosing components greatly. It would be fatal to the design of large complex system.

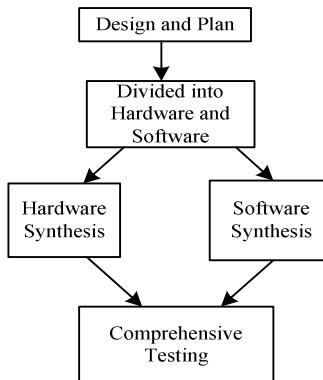


**Fig. 1.** Process of sequential design

### 2.2 Parallel Design of Hardware and Software

Fig.2 showed that this method began with the design and plan which used behavioral description to discover various ways to separate function modules and provided a basis confirm for the separation firstly. In the second step, the assessment of the

separation would help to find out the optimal choice accurately, which could verify and modify timely in early process of the design. The statues of hardware design and the software design were same in concept of parallel design, development of software need not wait for the completion of hardware, thereby it saved time, but this method didn't improve the constraints of embedded system. The developers must understand the structures and attributes when they chose the available IP modules, because of the black box essence of IP and restrictions of developers' knowledge and experience, it was impossible to understand all the suitable applications of IP, which also reduced the freedom of choice [4].

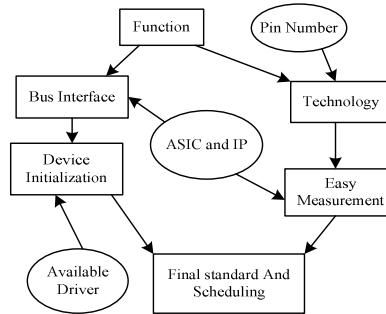


**Fig. 2.** The process of parallel design

Word processor, please use the font closest in appearance to Times. Avoid using bit-mapped fonts if possible. True-Type 1 or Open Type fonts are preferred. Please embed symbol fonts, as well, for math, etc.

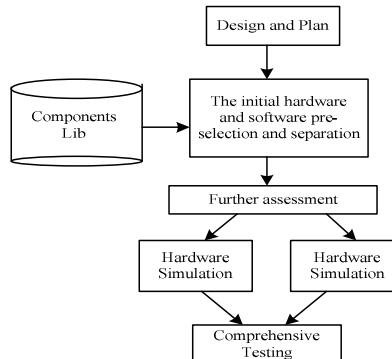
### 3 Parallel Design Based on Simulation

Parallel design based on simulation was the method that after the completion of the design and plan, the developers used the potential components to build system for the initial separation of software and hardware, components could be selected from the component repository, database or other information. The final result was choosing a set of hardware and software IP, the set of chosen components was the best programs that met the design standards. At first, considering all of the relevant factors in the process of the system design, removing ahead all of the possible errors, contradictions and conflicts. After that, the attention of evaluation and selection would focus on selecting standards such as easy measurement, which could be from the data sheets, technical manuals, and est. [5]. Establishing standards would be assessed further, and in this process attention would focus on specific standards for selection and scheduling of components like Fig.3 showed.



**Fig. 3.** Simulation Based Design Method

The most important standard to select suitable components was whether the Bus Interface matched with the microprocessor. In such situation, the shortcomings were big overhead, and there was significant latency between the communication of microprocessor and new components [2].

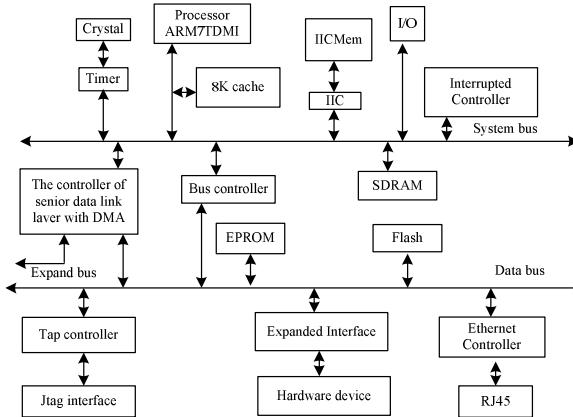


**Fig. 4.** Simulation based parallel design process

Now identification of the possible components completed, and components were arranged well. The method of simulation could assess system and identify the problems by real-time detection method, components that only passed this test could be used in the system, whose key was to build a simulation platform. The specific method was set in the next chapter. Fig.4 showed the process of the specific design.

#### 4 Simulation Platform Based on ARM

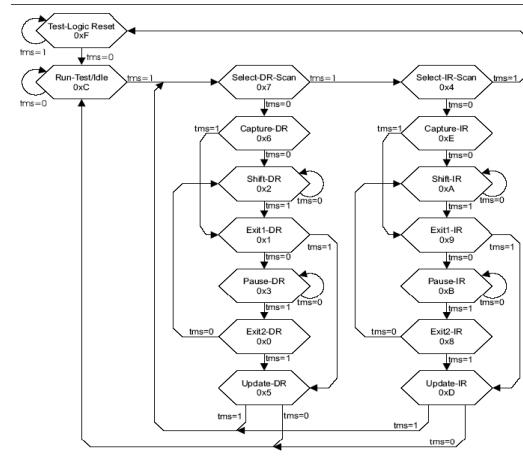
For the purpose to simulate and inspect the hardware and software components, the new method was to construct test environment as Fig.5, and ran the real-time simulation in that environment. For gaining the detail in the system of the complex embedded systems, the hardware and software were simulated and checked at the same time. When microprocessor was unable to complete a task, it would require additional hardware. Because in this situation ASIC would spend huge, so the best way was choosing the right components from the chips and IP [3].



**Fig. 5.** Basic structure of ARM simulation platform

#### 4.1 Hardware Simulation Supports

ARM7TDMI controlled the data bus by specific bus controller, which could realize the EPROM download easily. The simplest and the most common method was using JTAG emulator on simulation platform. Because JTAG debugged in the target platform, simulation was closest to the target hardware, many issues of interface such as restriction of high frequency operation, AC and DC parameter matching, and restrictions of cable length were minimized. (Please see IEEE 1149.1 standard about detail of JTAG parameters).



**Fig. 6.** TAP state conversion

Boundary scan in JTAG simulator was very important, the chip would provide several independent boundary scan chain to achieve functional testing. Boundary scan chain input and output serially, it would be convenient for observing and controlling

the chip in debug by the corresponding clock signal and control signal. The boundary scan chain was controlled by the TAP controller, including interface of input signal such as TCK, TMS, TDI, TDO and interface of output signal such as TRST. The process to visit the data registers by TAP was to select a needed visiting Data Register (DR) by the Instructions Register (IR); the needed data was input into the selected Data Register by driving TCK and TDI; at the same time the data in the selected Date Register was read out by TDO. The TAP [6] state machine was shown in Fig.6.

## 4.2 Software Simulation Supports

After integration of ARM7TDMI and Cache, SDRAM and Flash Memory, it provided interrupt controller and programmable I/O ports further, which enabled that the system had sufficient resources to run advanced software components and connected with the PC closely by UART, and realized that PC communicated with applications running on ARM7TDMI by COM port.

The simulation of software was the high level simulation model, the algorithm of behavior and the mixture of structure were described as object, which skipped the structure of circuit and the mainly focused on the functions of system and internal process of operation. The basic elements became operation and process mainly considering about data transmission, timing match, operation processes and state transition between each operation [8]. The method of high level simulation was to implement and explain the description, which was described directly by the high level language, or implementing C or C++ language that was compiled from HDL. The key of software simulation was how to deal with Parallel events, which must be managed by the independent time.

## 5 Conclusion

The basic idea of parallel embedded system structure design based on simulation was to get detail of internal systems operation in complex embedded systems, the difference between this design method and other major methods was simulation. And compared to simulation on the computer, this simulation method can check serious problems real-time when system operated, such as reducing response time and switching consumption of delay resource.

- 1) The need to achieve hardware configuration, performance analysis and optimization, code editing and compiling, running and debugging procedures, etc.
- 2) Many third-party components in different development platforms used different programming languages, which made systems heterogeneous
- 3) The realization and synergy implementation of the system when the components were in the network environment.

## References

1. He, J.-M.: Embedded 32-bit microprocessor system design and application. Electronic Industry Press, Beijing (2006)
2. Nitsch, C.: Design of Emulation-based Real-Time Embedded System,  
<http://www.fiz.de/>
3. Li, J.-G.: Detailed Explanation of ARM Application System Development. Tsinghua University Press, Beijing (2003)
4. Wolf, W.: Hardware-Software Co-Design of Embedded System. Proceedings of the IEEE (July 1994)
5. Wei, K.:  $\beta$ : Performance Analysis of Real-Time-Operation Systems by Emulation of an Embedded system. In: 10th IEEE Workshop on Rapid System Prototyping, Clearwater, Florida, USA (1999)
6. OPEN - JTAG development team: ARM JTAG Debugging principle (October 2004)
7. Lei, J., Ro, K.-L.: Embedded Simulation Development Environment Architecture. Journal of Electronic Science and Technology University 32(6) (2003)
8. Chen, D.-J.: Embedded Software Development System Simulation Study. Electronic Journals 28(6) (2000)

# **Research on the Object-Oriented Unit Testing Based on the Genetic Algorithm**

Kehong Zhang

School of Information Engineering  
Lanzhou University of Finance and Economics  
Lanzhou China  
hawkzhang1@163.com

**Abstract.** The object-oriented unit testing based on the genetic algorithm is a white-box testing method under the guidance of the demand analysis in software system. Traditional unit testing is mainly analyze the internal logical structure of products, but the object-oriented unit testing is not only considering the internal logical structure, but also mainly analyze the various features of class in the development of object-oriented. The paper combined with the genetic algorithm of artificial intelligence to testing the object-oriented software, so as to explore the new methods and techniques to improve testing efficiency.

**Keywords:** Genetic algorithm, Fitness, Object-oriented, Selection, Crossover, Mutation.

## **1 Introduction**

The guideline of unit testing is able to automatic generation the minimum program unit of compiled operation. This theory also can be used in software which is developed by object-oriented, but the shortage of the object-oriented testing is not clear to regard the method or class as a unit [3]. If the method as unit may cause confusion for object-oriented development and testing which is regarded the class as a unit. So it makes class as a unit testing in unit testing of the software system testing cases library.

Genetic algorithm is a calculation model which simulated biological evolution process of natural selection and genetic mechanism from Darwin's evolution theory. Its main feature is operating on the structure object directly, and with better ability of global optimization. In probability for optimization method, it can search space automatic acquisition and guiding optimization, and adaptive to adjust the search direction. It has been widely used in combinatorial optimization, machine learning, information processing, and adaptive control, etc. Based on the completeness of genetic algorithm apply in which regard class as a unit testing, but also the improvement of software testing technology.

## **2 Genetic Algorithm and Its Application in Software Testing**

Genetic algorithm is a better method for searching through simulating the natural evolution process, the essence of it is random search algorithm based on the

probability. Random set up the chromosome to be solved problems, each chromosome called population. And then testing the entire chromosome whether it can solve problem or not, and evaluate the ability of each chromosome for solving problems. If there are some chromosome can not solve, selecting two chromosomes as father generation in population according to the fitness of chromosome, genetic match to the next generation or through cross produce two offspring, then two generations make mutation with small probability, Repeat the process it get a group of offspring chromosome with the same number of father generation. With this group of the chromosome to testing, until we get a chromosome, it can solve the problem.

There are four steps of basic genetic algorithm. Firstly, produce random initial population. Secondly, calculate the fitness of every population member. Thirdly, estimate solution, if it has solution, the genetic algorithm will quit, and otherwise it will come into selection, crossover or mutation to produce new population of high fitness value. Last, get the optimization value and quit.

In this paper, using the genetic algorithm to realize the testing cases can be described as applied genetic algorithm to get a group of optimization testing cases. In the process of each optimization algorithm, the automatic generator of testing cases drives the tested procedures with the current population. Each execution path of testing cases will be tracked and recorded. Calculate the fitness value with instrumentation technology, and produce the next generation population. After the evolution of many generations, get the end of the optimization population or over the specific periods conditions.

### **3 The Sufficiency of Unit Testing in Software System Testing Cases Library**

Unit testing sufficiency is not only the important condition of class testing, but also the security of guarantee. During we design the unit testing cases of software system testing cases library, we should consider whether it can guarantee each sentence execute or not, whether it can find each Bug of program or not, at the same time, we also should consider the opposite problem, that is, whether it is necessary to find all the Bug or not based on the efficiency, and consider the features of class such as inheritance, polymorphism and encapsulation, these will cause some new problems that never meet at traditional testing. For doing object-oriented unit testing effectively will consider the adequacy of the class testing. There are three standards.

a. The adequacy based on the state transition of class. There are many kinds of class. It is necessary to consider the state transition during designing the testing cases. If the testing cases do not exhibit at least one state transition, it means a failure testing.

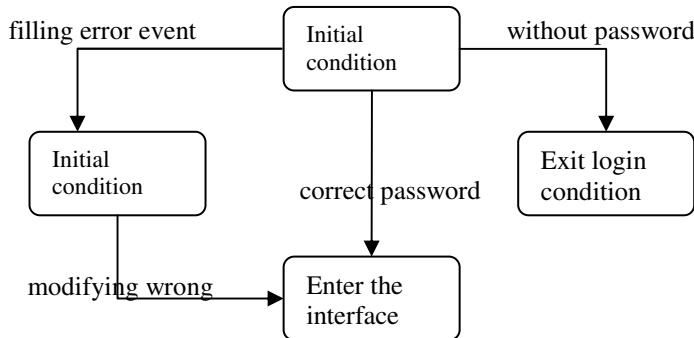
b. The adequacy based on limit. It may have precondition and postcondition in each operating. So we should consider the influence of these constraint conditions during designing testing cases.

c. The adequacy of paths. It also should consider the testing based on path whether is complete or not during designing the testing cases. In other words, after operate testing cases, each path of program at least perform once as much as possible. It mainly focuses on the efficiency of the testing in this paper.

## 4 The Designing of Object-Oriented Unit Testing with Genetic Algorithm

### 4.1 The Introduction of Example Program

The main tool of developing and designing is VC in software testing management system. The CTestingDlg class is used to login system. The class can cause three conditions through the initial state, and it needs 20 seconds to login the system. If it success to login, then enter into the interface. However, it needs to repeat the password when it is wrong. Its failure to enter after three times then can not use the system, and exit the login screen. It should be mentioned if the login time over 20 seconds, it also exits the system. The state transition as follows:



**Fig. 1.** The state transition login

### 4.2 The Analysis of Producing Testing Cases through Genetic Algorithm

1. Code rules: Because it is genetic algorithm, so the primary problem of generating testing cases is input variables into chromosome through code rules, so as to operate the genetic. That is due to genetic algorithm is focus on parameters code, so it will be measured a series of parameters of the tested program first, that is good for selection, crossover and mutation [1].
2. The choice of the fitness function: It is the key step. Because the structure of fitness function is very important. The choice of fitness function is mainly cause the effective of genetic algorithm, and also determines the algorithm whether can converge to the extreme value or not. It evaluates the solution molar. Adaptive function is defined different ways for different questions with experience, and constantly attempt.

In this paper we introduce “branch function instrumenter [4]” based on the features of concrete classes. That is, insert a real value function  $F(X)$  before each branch point designated by the logic path in the internal program unit. When a group of testing case driving is carried out by testing unit, these real value functions will be counted. The value of  $F(X)$  will reflect in the testing case, the deviation degree of actual execution path and designated logical path in tested unit. Combined with the detail condition of class, we find that void CTestingDlg::OnOK() function has three

milestones. Firstly, supposed time value is F1, and the maximum of time Max is 20 seconds, so set the recent time is time, then we get  $F1=time-MAX$ . Through the feature of function, the value of F1 is decided by the password whether it is correct or not and the length time of input. Secondly, the judgment of the correct password is set to F2, because it is wrong  $m\_recordset->GetRecordCount()==0$ , so we set a constant T, especially set the value of correct password to T, so  $F2 = i-T$ . Thirdly, the choice of times, because the maximum of input times is 3 in program, if over 3 times it will be illegal users , then close the program. So the fitness function  $F3=count$ . Therefore the whole fitness function  $F= F1 + F2 + F3$ , and we can get that the smaller value of F, the better it is.

3. Selection: The selection is the process of choosing an chromosome with strong vitality from the generation population to produce a new population. It will use the roulette wheel selection. That is, make each chromosome from the generation population form a circle roulette proportional through the reciprocal of fitness value F. Random turn the roulette. When it stops, the pointing chromosome is the selected chromosome. Because of the higher chromosome has wider area with the reciprocal of fitness value F, the selected probability is also higher. So it can guarantee the higher chromosome with the fitness value can produce more generations in a new population.

4. Crossover: The effectiveness of genetic algorithm is mainly from choice and crossover; especially it plays a key role in the genetic algorithm. It is realized by two steps. The first is random select two chromosomes from the population as the generation chromosome of intersection operation [2]. The second is random select crossover point, exchange the matching string, then it will produce new string. During the operation of crossover, crossover probability P is important to control parameters. It determines the update ability of chromosome and searching ability of algorithm in solution space. Therefore, this paper put forward the choice method of crossover probability P with heuristic information, that is, suppose to compare the matching string with password of database, the higher of the similarity, the greater value of probability P (As usual  $P<=0.8$ ). And then expand the digits of crossover according to the similarity, it also can reduce iteration times to make search quickly.

5. Mutation: Mutation algorithm is to change some gene value of chromosome, and maintain the diversity of population. At first, Mutation will select an chromosome random from the population, and then change the selected chromosome gene value of chromosome random with the probability. The same as nature, the occurrence probability of mutation in genetic algorithm is very low.

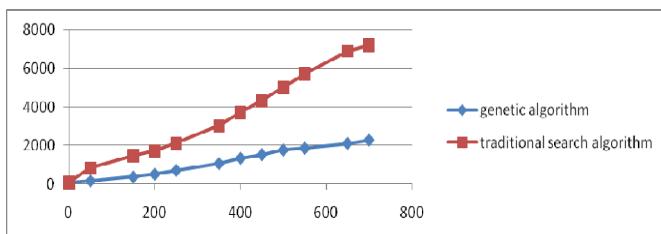
#### **4.3 Genetic Algorithm and Testing Results**

This paper defined the finishing symbol of genetic algorithm as follow: selecting the optimize chromosome, if it required each generation save the optimize chromosome, genetic algorithm is convergence. Each time it will produce new population, compare the optimal chromosome from new population in the way of fitness value. If the result is bad, the optimal chromosome from the generation population will instead of the others.

The genetic algorithm is:

1. Random produce n chromosomes for the initial population, and set evolution counter “count”, and the initial value of it is 0, the maximum is the maximum iteration.
2. Counting the fitness value of each chromosome in population according to the fitness function of instrumenter algorithm.
3. To choose the roulette wheel selection algorithm that is a classical one. Make all of the chromosomes in population choose the optimal chromosome through the reciprocal of fitness value F.
4. It has mentioned the crossover in this paper. Suppose to compare the matching with the password of database, if it has more similarities, the value of probability P is great. And use the way of expanding the digits of crossover according to the similarity.
5. Mutation is to change gene value of chromosome based on the small probability.
6. Compared the optimal chromosome of generation population with the fitness value of optimal chromosome from new population. The great fitness value of optimal chromosome will be selected in the new population.
7. As for population, return step 2. Until it get the satisfied selected standard or the iteration number “count” is max.

Through the above algorithms, we defined the value of P is 0.8. The result of detail testing as Figure 2 shows:



**Fig. 2.** Genetic algorithm delay and traditional search algorithm delay

## 5 Conclusion

In Through the survival periods of software, unit testing is white box testing, it is mainly aimed at program logic testing in the testing technology. On the other hand, object-oriented testing of object-oriented development technology is also different from the traditional testing. Especially the inheriting, encapsulation, and diversity of object-oriented classes caused great difficulties during the testing. So, in this paper combined with the genetic algorithm of artificial intelligence to design testing technology is the innovation and attempt. Therefore, the ideal practice is to strict censorship the formation document respectively of each stage according to the software engineering, and then put forward a concrete testing design. Anyhow we should improve robustness, correctness and effectiveness of software system testing cases library and testing efficiency through the various methods and new technology.

## References

1. Sun, J.-H., Jiang, S.-J.: An approach to generate test data for multi-path coverage by genetic algorithm. *Journal of Microminiature Computer Information* 2 (2010)
2. He, J.-W., Song, C.X., Liu, H.: Design and Implementation of Eight Puzzle Problem Based on Genetic Algorithms. *Journal of Computer Technology and Development* (March 2010)
3. McQuillan, J.A., Power, J.F.: A survey of UML- based coverage criteria for software testing. *National University of Ireland, Ireland* (2005)
4. Cao, X.-Y.: Study of automatic test data generation based on hybrid genetic algorithm. *Journal of Computer Engineering and Design* 31 (2010)
5. Priestley, M.: *Practical Object-Oriented Design with UML*, 2nd edn. Tsinghua University Press (May 2005)
6. Srivastava, P.R.: Optimisation of software testing using Genetic Algorithms. *International Journal of Artificial Intelligence and Soft Computing* (2) (2009)
7. Ahmed, M.A., Hermadi, I.: GA-based multiple paths test data generators. *Computers & Operations Research* (1) (2008)
8. Yao, Y.: New Test Case Generation Method Based on Genetic Algorithm. *Journal of Computer & Digital Engineering* 1 (2009)

# Data Aggregation and Information Type in Road Probing

Lin Sun, Ying Wu, Jingdong Xu, Jinchao Li, and Yuwei Xu

College of Information Technical Science, Nankai University,

Weijin Road 94, 300071, Tianjin, China

{snova,lijincha,sky.love}@mail.nankai.edu.cn,

{wuying,xujd}@nankai.edu.cn

**Abstract.** The RSU can collect and disseminate information to enhance road traffic safety. Meanwhile, with the widely use of the sensors and wireless ad hoc networks, the VSN is formed. Combined the functions of the RSU and the VSN, the Road Probing scenario is proposed to enlarge the RSU's information collection range. In this paper, we focus on the safety-related traffic information on freeway and the practicable data aggregation scheme for Road Probing. We conclude the most important data types to be gathered. After that, a data aggregation scheme using time converge globally and space converge locally will be proposed in Road Probing. Our method can greatly reduce data redundancy, and its advantage is confirmed by the simulation results. The data reception improves greatly and the transmission delay is comparatively low.

**Keywords:** VANET, data aggregation, RSU, Road Probing.

## 1 Introduction

The Vehicular Sensor Network (VSN) is a novel application network, which combines the Vehicular Ad Hoc Network (VANET) and the sensors. In VSN, each vehicle is a sensor to collect the environmental information as well as a wireless node to transmit the collected data to the central equipment. Road Probing is a new VSN application scenario proposed in [1] with the use of the Road Side Unit (RSU). In that scenario, the RSU will launch the probing process and select some vehicles as probes. The probes complete data acquisition and dissemination together with the RSU. The data can be used to inform the drivers of road status and potential dangers on the freeway. This paper mainly focuses on the safety message collection and dissemination. Firstly, we introduce the application scenario and give two probe selection models. Secondly, we focus on road traffic information and discuss practical data types in Road Probing. The data can be sorted into four categories to truly reflect safety-related traffic situations. Furthermore, we analyze the importance of data aggregation in each probe, and propose our data aggregation method. Our method adopts both time converge and space converge schemes. It can improve the RSU's data reception rate and data processing efficiency. Finally, we implement the simulation program using NS tools and the public available freeway patterns to further validate our scheme. The data reception rates of two models using aggregation have at most 27.7% and 41.2% improvements. And packet transmission delays are less than 2.96 seconds in two models.

The remainder of the paper is organized as follows. Section 2 provides related work about Road Probing and data aggregation in VANET. We analyze the importance of data aggregation and present the information types of road traffic as well as our data aggregation method in Section 3. Section 4 presents the simulation results and the performance evaluations. The conclusions are given in section 5.

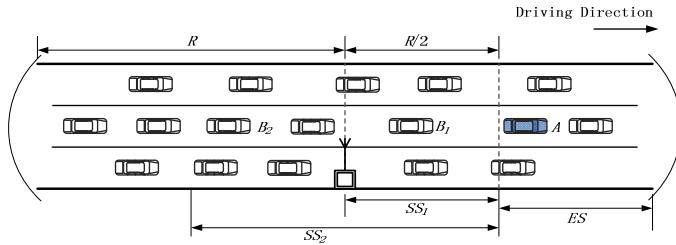
## 2 Related Work

Road Probing can automatically collect traffic and environmental information with the cooperation of vehicles and the RSU. The probing information can provide drivers cruise-assist services [2]. Paper [3] further develops the probe data to detect incidents in highway systems. The function of data aggregation is fusing information to reduce redundancy and data traffic in each node. Paper [4] first introduces the concept of VSN and builds the data aggregation application framework. Catch-up [5] is an aggregation scheme, and its basic idea is adaptively changing the forwarding delay of individual reports in a manner that a report can have a better chance to meet other reports. The Catch-up scheme is a tradeoff between communication overhead and propagation delay. Focusing on the data aggregation and the RSU placement for a VANET traffic information system, paper [6] proposes a genetic algorithm which is able to identify good positions for static RSUs in order to cope with the highly partitioned nature of a VANET in an early deployment stage. However, as with Catch-up, this work does not disseminate or aggregate information for safety applications. Using compression to provide aggregation without losing accuracy, CASCADE [7] is a cluster-based aggregation scheme, and is improved in [8] to reduce bandwidth consumption and extend driver visibility.

## 3 Data Aggregation in Road Probing

### 3.1 Road Probing Scenario and Probe Selection

Road Probing is an application scenario with the characteristics of VSN, in which the RSU chooses some vehicles as probes to collect traffic information. We assume that all the vehicles and the RSU are equipped with wireless devices to enable their communications. The communication range is denoted as  $R$ . The selected probe has the responsibilities of gathering local information, receiving probing packets from the previous probe, doing data aggregation, and broadcasting the aggregated packets. Each probe is a data collector as well as a data forwarder. Therefore, a one-dimensional and multi-hop chain structure network is formed. Fig. 1 depicts probe selection procedure. One quarter of the RSU's coverage located on the right side is reserved as Exit Section ( $ES$ ). When the previous probe (e.g. vehicle A) enters  $ES$ , the RSU starts to choose the next one. There are two models to follow: 1) Model 1. Select the probe in Select Section 2 ( $SS_2$ ). The distance between two consecutive probes is no longer than  $R$ . Model 1 can minimize the number of probes to simplify the transmitting procedure and reduce the packet delay. 2) Model 2. Select the probe in Select Section 1 ( $SS_1$ ). The distance between two consecutive probes is no longer than  $R/2$ . Model 2 will add redundant nodes to make the chain more robust and increase the RSU's reception rate.



**Fig. 1.** The selection of probes

When the vehicles drive into  $SS_1$  or  $SS_2$ , the RSU will receive their beacons and keep the vehicle's ID, location and driving lane into a list. The list will be updated at all time. When the previous probe enters  $ES$ , it will send a notification to the RSU. After receiving the notification, the RSU randomly chooses a car on the middle lane from the list as the next probe. Especially, if the notification is lost, the RSU still can choose the next probe after a specific period. Then the RSU will inform it using broadcast, and allocate a probe ID to it. The new probe will update its status and starts to receive the probing packet and forward it after data aggregation.

### 3.2 Data Type and Data Aggregation Protocol in Road Probing

One of the most important functions of VANET is to avoid the traffic accident and improve traffic security. We conclude four kinds of safety-related information: 1) data status, including time and location; 2) vehicle status, including driving direction, speed and acceleration; 3) road status, including smoothness and visibility of the road and 4) special status, including whether to have an accident and so on. The objective of Road Probing is to collect information and use it to improve the traffic security. So the RSU has to collect all the information types. The representative data items to be handled in our data aggregation scheme are timestamp, location, vehicle speed, road visibility and whether to have an accident. Those contents can be mutually complementary to periodically reflect traffic conditions. Table 1 shows the information and corresponding types. Those items in Table 1 can be designed in four different data structures: 1) time value, 2) place value, 3) average value and 4) special value. Timestamp is a time value to present whether the received packets belong to a specific period. Location is a place value to indicate whether the packets belong to a specific area. Average value includes vehicle speed and road visibility. The two items represent distinct traffic conditions, yet in the aggregation they have the same way to process - averaging. As the initiator of Road Probing, the RSU concerns about the traffic condition of the entire freeway other than individual vehicles. Therefore, though averaging data will lose some accuracy, they can reflect overall situation. The last type corresponds to the parameter "whether to have an accident". In our scheme, it only has two values: 1 indicates accident and 0 indicates none. Combined with different space areas, the parameter will be expressed as a string of 0 and 1.

**Table 1.** The information and corresponding type

Data Type	Information
Data status	Timestamp, Location
Vehicle status	Vehicle speed
Road status	Road visibility
Special status	Whether to have an accident

Data aggregation is rather important in Road Probing scenario. If it is not used, the probe will individually transmit local probing packet and the one from the previous probe. Denote the number of probes in the chain as  $n$ , the  $m^{\text{th}}$  probe ( $1 \leq m \leq n$ ) needs to broadcast  $m$  probing packets in a time slot. During one slot, there will be  $n$  packets totally. Denote the bandwidth consumption as  $B$  and the packet length as  $PL$ . The value of  $B$  can be computed as  $PL \times n$ . If data aggregation is used, the probe will diffuse local packets and the forwarded ones to generate a new one. Then each probe only needs to broadcast 1 probing packet in a time slot, and total number in the system is also 1. Thus, the bandwidth consumption  $B'$  can be calculated as  $PL \times 1$ . In ideal situation, the saved bandwidth  $B_s$  will be  $PL \times (n - 1)$ . As we can see, data aggregation scheme can greatly reduce the packet number to save wireless bandwidth and lower packet collision rate. Moreover, if data aggregation is not adopted, the RSU will receive and process lots of packets with redundant information, which greatly improves the RSU's working load and reduces its working efficiency. Data aggregation can be divided into time converge and space converge. Time converge is to diffuse the packets based on their timestamps and new timestamp is the sum of the old ones. Space converge diffuses the packets based on their locations. In our scheme, each probe will gather packets in its transmission range and use space converge to generate a new probing packet. In addition, each probe will do time converge after receiving the probing packet, and the location of the packet will not change.

## 4 Simulation Experiment

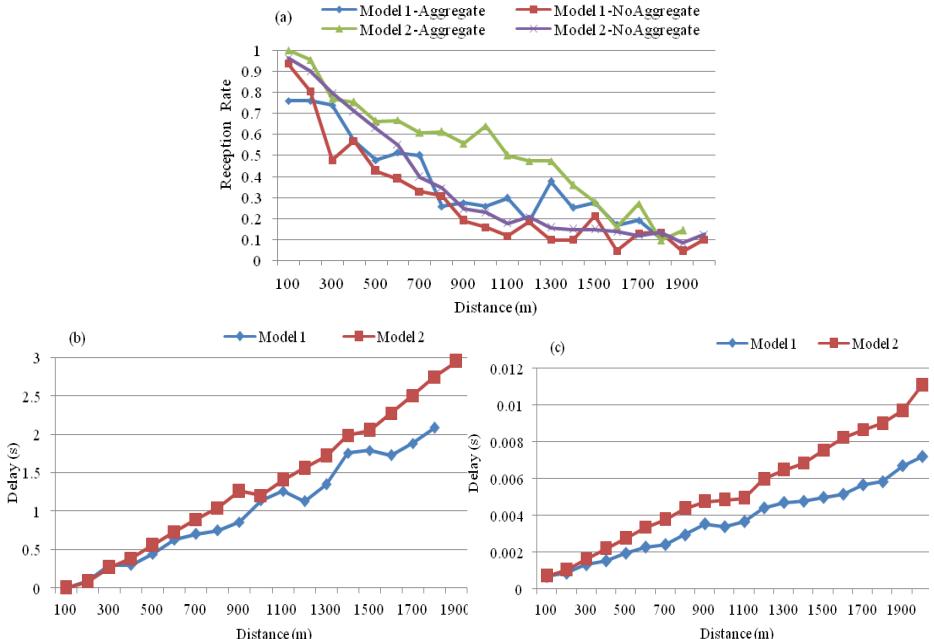
### 4.1 Simulation Setup

We build a simulation environment for Road Probing on ns-2.34, and employ the public available freeway patterns and ns traffic trace generation tools presented in [9] to obtain a realistic scenario with 60 seconds simulation time. The simulated freeway is 12km long composed by 3 lanes in one direction. The vehicle number is 467 with the average speed of 120 km/h and a fixed RSU is placed at the beginning of the 6<sup>th</sup> kilometer. The traffic density is 11 vehicles per lane per kilometer. As suggested in [10], we use the probabilistic wireless propagation model Nakagami with the fading intensity 3 in the simulation. The PHY and MAC layer parameters are carefully configured according to IEEE 802.11p protocols. The communication radius  $R$  is around 300 meters with our configuration. According to DSRC protocol [11], we choose 10 pps (packet-per-second) as the beacon rate and 2 pps as the probing

message rate. The data length of each message is set to 300 bytes, which is the approximate middle value of reasonable safety message size for VANETs.

## 4.2 Performance Evaluation

In our simulation, we compare the RSU's reception rate and transmission delay with and without data aggregation. Fig. 2 (a) indicates the metric of the RSU's reception rate with respect to the distance to the RSU from where the probing packet is broadcast. As the figure shows, the reception rate with aggregation is much higher than the rate without it. The improvements are at most 27.7% in Model 1 and 41.2% in Model 2. We can also find out the difference between the rate with and without aggregation in Model 2 is greater than that in Model 1. Because there are fewer probes and probing packets in Model 1, and the packet loss has greater effect upon the reception rate. Fig. 2 (b) and (c) are the transmission delay comparisons with respect to the distance from the probe to the RSU. As we can see, the transmission delay with aggregation is much higher. In our scheme, the probes broadcast safety messages every 0.5 second, so the probing packets forwarded from one probe to the next have to wait 0.5 second at most, which causes relatively long transmission delay. However, the longest delay is no more than 2.96 seconds, which is still an acceptable value. As shown in the figure, the delay in Model 2 is slightly higher than that in Model 1. The reason is that Model 2 has more probes, and the probing packets go through more relay nodes and the transmission time is raised accordingly.



**Fig. 2.** a) The RSU's data reception rate; b) Transmission delay with data aggregation; c) Transmission delay without data aggregation

## 5 Conclusions

We have analyzed the traffic information type and defined the data structure. More importantly, data aggregation is critical to save wireless bandwidth. We have proposed an information-based data aggregation scheme in Road Probing scenario. With great practical significance, our scheme can greatly improve data reception rate and the RSU's working efficiency. Our future work will focus on probe selection strategy and connectivity recovery method to enhance information collection.

## References

1. Yang, L., Xu, J., Wu, G., Guo, J.: Road Probing: RSU Assisted Data Collection in Vehicular Networks. In: Proc. WiCOM, Beijing (2009)
2. Yamada, H., Makino, H., Takamune, M., Wakamiya, M., Takenaka, K., Nomoto, T., Inoue, H.: Cruise-Assist Services Utilizing Up-link Information. In: Proceedings of the 12th World Congress on ITS, San Francisco (2005)
3. Hirai, S., Hatakenaka, H., Watanabe, Y., Ogane, K., Kojima, M.: Incident Detection by Probe Data. In: Proceedings of the 13th World Congress on ITS, London (2006)
4. Lee, U., Magistretti, E., Zhou, B., Gerla, M., Bellavista, P., Corradi, A.: Mobeyes: smart mobs for urban monitoring with a vehicular sensor network. *IEEE Wireless Communications* 13(5) (2006)
5. Yu, B., Gong, J., Xu, C.Z.: Catch-Up: A Data Aggregation Scheme for VANETs. In: Proceedings of ACM VANET, San Francisco (2008)
6. Lochert, C., Scheuermann, B., Wewetzer, C., Luebke, A., Mauve, M.: Data Aggregation and Roadside Unit Placement for a VANET Traffic Information System. In: Proceedings of ACM VANET, San Francisco (2008)
7. Ibrahim, K., Weigle, M.C.: CASCADE: Cluster-Based Accurate Syntactic Compression of Aggregated Data in VANETs. In: GLOBECOM Workshops (2008)
8. Ibrahim, K., Weigle, M.C.: Optimizing CASCADE Data Aggregation for VANETs. In: 5th IEEE International Conference on Mobile Ad Hoc and Sensor Systems, Atlanta (2008)
9. FuSler, H., Torrent-Moreno, M., Kruger, R., Transier, M., Hartenstein, H., Effelsberg, W.: Studying Vehicle Movements on Highways and Their Impact on Ad-hoc Connectivity. Department of Computer Science, University of Mannheim, Tech. Rep. TR-2005-03 (2005)
10. Taliwal, V., Jiang, D., Mangold, H., Chen, C., Sengupta, R.: Empirical Determination of Channel Characteristics for DSRC Vehicle-to-vehicle Communication. In: Proceedings ACM VANET, Philadelphia (2004)
11. White Paper: DSRC Technology and the DSRC Industry Consortium (DIC) Prototype Team (2005)

# The Smart Card Remote Unlocking Method and Its Implementation

Yongtao Hu<sup>1</sup>, Xing Wang<sup>1</sup>, and Yunlu Gao<sup>2</sup>

<sup>1</sup> Key Laboratory of Information Network Security,  
Ministry of Public Security  
201204 Shanghai, China  
[sweetjohnhu@163.com](mailto:sweetjohnhu@163.com)

<sup>2</sup> School of Software Shanghai Jiao Tong University  
201204 Shanghai, China

**Abstract.** In order to protect smart card from being misused and leaking out sensitive data, security mechanisms of locking after repeatedly retry will be introduced while designing chip operating system, with which the smart card will be automatically locked if it is threatened by illegal operations. This mechanism can not only prevent brute-force attack, but also trouble the users, for unlocking method should be provided to legitimate users after attacks or misuses. The security of traditional unlocking methods depends on the shared unlocking key between smart card and unlocking tool; once the unlocking tool is lost or comprised, security risks will be caused. The remote unlocking method proposed by this paper unifies the unlocking interfaces, manages the unlocking keys centrally and adopts a multiple-level and flexible model containing key management center, agent and users. This can significantly reduce the risks of key leakage and loss of control, and also has the features of controlling the unlocking operations, tracing operating staffs, auditing total usage and deploying flexibly.

**Keywords:** smart card, security mechanism, unlock, key management.

## 1 Introduction

While designing chip operating system for smart card [1-3], access control mechanism will be required for critical operations in order to protect the card from being misused and to prevent sensitive information leaking. It's common for card-holders to input PIN before critical operations, which are only granted after the success of PIN verification [4]. The number of attempts is limited to stop brute-force attacks. Thus after a certain number of attempts, the card will be automatically locked and sensitive information will not be exported any more [5]. If there is no unlocking mechanism correspondingly, card-holders have to initiate the card and download the user certificate again, which involves a series of complicated operations, such as revocation of former certificates, and will causes great inconveniences to users.

When the card is locked as the security mechanism plans, it could generally be unlocked with special unlocking tools [6]. The unlocking principle is that when the

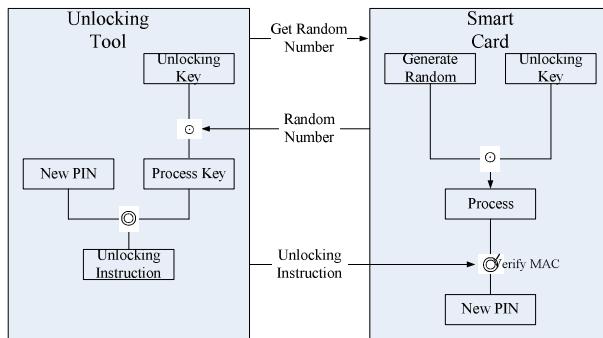
card is produced or issued, it usually pre-shares an unlocking key that is used to check whether the unlocking instruction is from the legal institution holding the same unlocking key. Unlocking tools commonly generate unlocking instructions according to the unlocking keys, and send it to smart card to unlock the PIN as needed. The general process is as following:

- Unlocking tool gets a random number R from smart card;
- Unlocking tool computes process key RK=C(R , Key) by unlocking key and the random number.
- Unlocking tool encrypts the new PIN with RK and calculate its MAC, generating the unlocking data:

$$\text{Data} = C(\text{newPIN} , \text{RK}) + \text{MAC}(C(\text{newPIN} , \text{RK}), \text{RK});$$

- Unlocking tool sends the unlocking instruction and the data to smart card;
- Smart card also computes a RK by the latest generated random number;
- Verify the data by RK. If success, then decrypt the new PIN and unlock the smart card.

The whole process is shown as figure 1.



**Fig. 1.** The process of traditional unlocking method

In the whole unlocking process, the unlocking key doesn't expose directly in the communication process. Besides, the process key is actually one-time padding because of the random number, so it ought to be secure. But there are still some shortcomings of this approach:

- A. *Unlocking operations of the card are uncontrollable.* Once the unlocking tool is controlled, unlocking operations can be applied to any card, and the using scope of the unlocking tool is not controllable and limited.
- B. *Unlocking operations of the card cannot be audited and tracked.* For the smart card can only communicate with the unlocking tool and lack of sufficient storage space, we can't audit the unlocking operations and also cannot determine the staff unlocking the cards.
- C. *The risk of unlocking key leakage is enlarged.* The reason that unlocking tool can reset the security state of the card is that there is a pre-shared unlocking key built-in the tool itself or stored in independent devices. Once the unlocking tool containing the pre-shared key is decompiled or cracked, the security structure of the smart card will be at high risks.

It is more secure to unlock the card using special tools by specified staff in specified places, but it's not convenient, for it causes extra burden to users. In the view of these considerations, we design a multi-level, extensible and secure unlocking model containing key management center, agents and users. It performs the unlocking operations as a service, and users can request unlocking instructions from key management center through the network. All the operations are required to be authenticated and audited, which not only improves the security, but also ensure that the services are widely available.

## 2 Smart Card Remote Unlocking Approach

Considering the defects of current techniques, we propose an approach of unlocking smart card remotely. It is a multiple-level, extensible model containing authority center, agent and users; unlocking instructions are encrypted in the whole process and one-time padding; the core unlocking keys and interfaces are centrally managed.

The remote unlocking system involves the following components:

- Key management center. It's responsible for maintaining the life cycle of the keys, including storage, usage, destroy, backup and recovery as needed. And it's identified by the root certificate.
- Authority and audit component. It is responsible for verifying the request from agents and auditing according to the set rules.
- Agent of all levels. They are in charge of verifying the requests of the next-level agents or card-holders, and authorizing level by level.
- Client software for unlocking. It ought to generate the unlocking requests, submit data for identity verification, receive single-valid unlocking instruction and perform the unlocking operations.

The whole unlocking process is as following:

- The user looks for the agent nearby to submit unlocking requests;
- Agents check whether the submitted information is legal or not; after the successful verification, client software is started and users are required to input a new PIN. Client software obtains a random number R from smart card and encrypts it together with new PIN by root certificate of key management center and sends it to the agent of upper level as unlocking request data RData and requests unlocking services.
- Agents forward the request to upper level until the authorization and audit module.
- Authorization and audit module verifies the identity of the requesting agent, and requests unlocking service from key management center if success.
- Key management center decrypts the request data RData with private key, getting the random number R and the new PIN. Then compute the process key  $RK=C(R,Key)$  by unlocking key Key and the received random number.
- Unlocking tool encrypts the new PIN with RK and calculate its MAC, obtaining the unlocking instruction data,

$$\text{Data} = C(\text{newPIN}, RK) + \text{MAC}(C(\text{newPIN}, RK), RK);$$

- Unlocking tool sends the unlocking instruction and related data to smart card;
- Smart card also calculates a RK with the latest generated random number,  $RK=C(R,Key)$ ;
- Verify Data with RK. If correct, then decrypt the new PIN and unlock the smart card.

Compared with traditional approaches, this proposal has the following advantages:

- Unlocking key doesn't appear in the domain of users and authorities and is only stored in key management center.
- Key management center provides public unlocking service through the Internet, making the unlocking services widely available.
- The identity of card-holders is verified by agents, and is forwarded to key management center when applying unlocking services.
- Unlocking services can only be requested and used by agents, whose identities can be verified using PKI.
- Key management can pre-install different unlocking keys according to different types of cards, thus, even if the unlocking key is leaked, only parts of the cards are affected, not the whole smart card system.
- Implement as above, all the operations are completely logged and is easy for auditing or tracking later on.

### 3 Implementation of Key Technologies

#### 3.1 Key Management Center

In order to protect the security of the unlocking key in its life cycle, key management center needs dedicated cryptographic equipment to securely store, use, destroy, backup and recover the unlocking key as needed.

- 1) *Storage of the keys.* Keys are stored in dedicated cryptographic equipment, making sure that they cannot be imported and can resist physics attack in some degree. When the devices are physically damaged, keys are automatically destroyed.
- 2) *Usage of the keys.* After loading the keys, the dedicated cryptographic equipment provides services to authority center through special service interface and finds the key to be used by the identification information in the unlocking request.
- 3) *Destruction of the keys.* The dedicated devices must ensure that the destroyed keys cannot be retrieved through either logical measures or physic methods, and all the backups are destroyed at the same time.
- 4) *Backup of the keys.* To avoid the unavailability of the service caused by failures of cryptographic equipment, the keys should be backed up in customized cryptographic equipment after being imported, which is offline and securely placed after importing the keys.
- 5) *Recovery of the keys.* When the keys in dedicated cryptographic equipment become invalid or damaged, they can be recovered from the backups.

### 3.2 Authorization and Audit Module

- 1) *Identification and Authentication.* In order for the security of the smart card system, unlocking service is only provided to the authenticated agents. Identification module enrolls the inferior agents, records the business information including the digital certificates, and makes service policy accordingly. When receiving the requests from the agents, the module firstly identifies the requestor by digital signature, secondly checks the integrity and consistency of the data, and finally authorizes according to the policy.
- 2) *Audit.* All business data in authority center is auditable and open, and can be taken use of to revise the unlocking policy.

### 3.3 Agents

- 1) *Identification and Authentication.* To facilitate the hierarchical and extensible deployment, other authentication and authority methods are provided besides the ones provided by authority center. Extended interfaces of identity management include:
  - Registration, record identification information.
  - Identification, implement the interfaces of identification and verification.
  - Authorization, authorize according to the authentication.
- 2) *Audit.* Unlocking services from agents of all levels provides not only the same authentication and authorization method with authority center, but also the function of tracking the specific operating staff.

### 3.4 Client Component

- 1) *Generate unlocking request.* Gather the basic information including series number, software and hardware version number, and then compose the request along with other request data produced by the card.
- 2) *Receive unlocking instructions.* Check the integrity and consistency of the unlocking instruction.
- 3) *Unlock the card.* Send the single-valid unlocking instruction to the card and complete the operation.

## 4 Evaluation of the Security

The security of this unlocking method lies in the following facts.

- In order to make sure of the security of PIN-unlocking process, the following two points must be guaranteed:
  - a) Unlocking instructions are not obtained direct through unlock key calculation;
  - b) The random number generated by the smart card makes the process key that actually encrypts unlocking instructions one-time padding.

- It can effectively reduce security risks, such as key leakage and loss of control, that challenge traditional unlocking services and methods. By storing the core key in dedicated devices and managing the life cycle of the core unlocking key through key management center, we can promote the security of the keys, and reduce the possibility of key leakage and loss of control.
- Unlocking operations are controllable, and operators can be tracked. Through centralized authentication and authorization, unlocking operations can be effectively managed and controlled, and the loss of control of unlocking services can be avoided. At the same time, rigid authentication mechanism can ensure that each unlocking operation will be traceable and the operating staff can be determined.

## 5 Conclusion

After analyzing the security defects of current techniques of unlocking smart card, this paper proposes a remote unlocking method based on actual needs and presents the implementation of some key techniques. The security analysis manifests that, compared with previous unlocking methods, this one can effectively reduce the risks of key leakage and loss of control; unlocking operations can be controlled, and operators can be tracked. Besides, the total usage is auditable and the deployment is flexible and extensible.

## References

1. Ravi, S., Raghunathan, A., Kocher, P., et al.: Security in Embedded Systems:Design Challenges. In: Transactions on Embedded Computing Systems, New York, pp. 461–491 (2004)
2. Kocher, P., Lee, R., McGraw, G., et al.: Security as a New Dimension in Embedded System Design. In: Proceedings of the 41st Annual Conference on Design Automation, pp. 753–760. ACM Press, New York (2004)
3. Jean-Francois, D., Nathalie, F.: Present andfuture smart cards (EB/OL),  
<http://www.it-c.dlk/courses/DSK/F2003/smart2.pdf>
4. Rankl, W.: Smart Card Handbook, 2nd edn. Wiley & Sons (2000) ISBN 0471988758
5. NIST, The Federal Information Processing Standard (FIPS) Publication 140-2 (EB/OL),  
<http://csrc.nist.gov/publications/fips/fips140-2/fips1402.pdf>
6. Smart Card Alliance. Privacy and secure identification systems: The role of smart cards as a privacy-enabling technology. A Smart Card Alliance White Paper (February 2003),  
<http://www.smartcardalliance.org/>

# A Rapid Review Method to Apply Digital Certificate Based on CA-Trust Set

Jingjing Yao<sup>1</sup>, Yongtao Hu<sup>1</sup>, and Yunlu Gao<sup>2</sup>

<sup>1</sup> Key Laboratory of Information Network Security, TRIMPS  
Shanghai, China

<sup>2</sup> School of Software, Shanghai JiaoTong University  
Shanghai, China

{Ningzifeng, Gaoyunlu711}@gmail.com, sweetjohn@163.com

**Abstract.** In the information society where digital certificates are widely used, users probably own more than one certificate. Generally, users are required to apply certificates on-site so that RA can verify the authenticity of the information submitted by users. Online application can improve the efficiency of RA, but it is still necessary for RA staffs to manually check the information submitted online. In order to realize automated verification and improve the efficiency of RA further, a CA-trust set is designed for RA, of which every element is CA and is approved by the current RA. Users add a signature to the information submitted to RA with a digital certificate issued by some CA belonging to a CA-trust set, so that RA can complete the verification of information through checking the signature.

**Keywords:** CA-Trust Set, RA, trust model.

## 1 Background

PKI is a key management platform that follows established standards, providing encryption and signature services. It manages the public keys in terms of certificates which binds the users' public keys to user information (such as names, Emails, ID numbers and so on) together and stores them in the certificates [1]. PKI is typically composed by five parts: certificate applicants, registration authority, certificate authority, certificate repository and relying party. Among these five parts, registration authority provides the following functions: submitting certificate application, verifying certificate application, submitting revocation application, verifying revocation application, submitting recovery application, verifying recovery application, issuing verification results, inquiry users, look over user certificate, delete users and so on.

The credibility of user identity is the stepping stone that determine whether the user can be granted or not, and is also the fundamental of the PKI system security [2]. As a result, verification of user identities is an important job of RA. User information is submitted when the users are applying for certificates. There are two methods of certificate application: online application and on-site application:

**On-Site Application:** users go to RA directly and fill out related forms. Then staffs of RA verify the information and submit to RA servers.

**Online Application:** users submit the information to RA through the Internet. RA verifies the information according to designed policies, and determines whether to accept the application or not.

On-site application enables RA staffs to manually check the authenticity of the information, but is not convenient for users. Online application simplifies the users' work. However, RA still has to spend human resources to complete the verification. Moreover, while verifying the information, RA cannot compare and confirm the information as conveniently as on-site application. Therefore, both approaches counter the same problem: low efficiency in information verification. And this becomes the bottleneck for increasing requirements of issuing bulks of certificates.

Actually, in the information society where digital certificates are widely used, users may already own one or more certificates issued by other CA before applying for a new one. For example, a user may own digital certificates of several banks for account operations at the same time. When applying for later certificates, RAs of corresponding PKI actually have completed the verification of user information, that is to say, the user information storing in the existing certificates is already be confirmed to be authentic and reliable. Thus, can this authentic information be used to verify the currently submitted application information?

This paper defines a CA-trust set for a given RA, and proposes a working process for users to apply online and for RA to quickly verify the application, which is a solution for current problem.

## 2 Fast Approval Process

### 2.1 Trust Structure

Different companies often have their own PKI architecture, and these different PKI systems are interrelated in practice. Service providers and recipients need to trust each other with a unified platform to establish collaborative relationships. There are some common trust models: Single-CA trust model, strictly hierarchical trust model, network trust model, the bridge CA trust model, user-centric trust model, etc [3]. No matter what kind of trust model is, it should realize PKI interoperation. Application interoperability [3] refers to cross-domain users and applications obtain information from different PKI system on a variety of security services, and complete a variety of applications. Interoperability is related to the transaction object participating in interoperability [5]. For the same thinking, a lightweight trust structure based on CA-trust set is defined as following to simplify the review process of RA. This structure need not have the complete characteristics of a typical trust model, but be responsible only to the specific application.

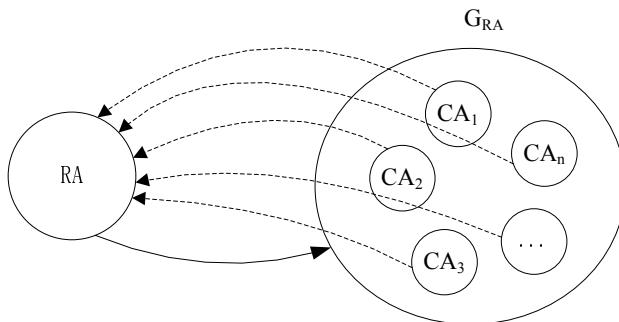
## 2.2 CA-Trust Set

The precondition of this application is that the current applicants already own several certificates issued by other CA, and certificates issued by this CA are all approved by the current RA. The concept of CA-Trust Set for a given RA is defined as below:

$G_{RA}$  is a set of CA:  $G_{RA} = \{CA_1, CA_2, \dots, CA_n\}$ ,  $n \in N$ .  $G_{RA}$  is regarded as a CA-Trust Set for RA if its element denoted as  $CA_i$  satisfies two conditions:

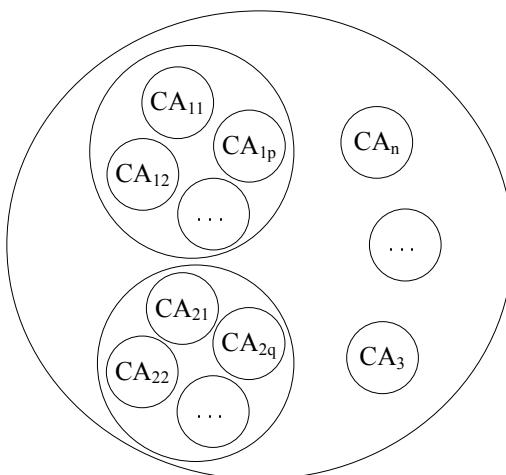
- 1) RA accepts all the certificates issued by  $CA_i$ ;
- 2) RA itself owns a certificates issued by  $CA_i$ .

That's what Fig.1 shows:



**Fig. 1.** The definition of CA-Trust Set

In Fig.1, a dashed arrow from  $CA_i$  to RA means that RA is verified by  $CA_i$  and RA own a digital certificate issued by  $CA_i$ . A solid arrow from RA to  $CA_i$  means that RA approves all digital certificates issued by  $CA_i$ .



**Fig. 2.** A special case of CA-Trust Set

As a special case of a CA-Trust Set for RA,  $CA_i$  could also be a set of CA, which is showed by Fig.2:  $CA_i = \{CA_{i1}, CA_{i2}, \dots, CA_{im}\}$ ,  $m \in N$ . In this case, single element of  $CA_i$  cannot establish a trust relationship between RA and itself. Only when every  $CA_{ij} (j \in [1, m])$  satisfies two condition proposed before and users own certificates issued by every  $CA_{ij} (j \in [1, m])$ ,  $CA_i$  can establish a trust relationship between RA and itself. The following discussion will use the base case showed by Fig.1.

To be used in approval process, a CA-Trust Set should satisfies another condition: certificates issued by CA should have something in common with the ones issued by  $CA_i$  in  $G_{RA}$ , such as similar verification policies of certificate, or similar private key information items of users' stored in certificate (names, ID number and e.g.), and so on.

### 2.3 Approval Process

For ease of description, mark the current user's certificates issued by  $CA_i$  belonging to  $G_{RA}$  as  $Cert_i$ , and corresponding public key and private key are noted as  $PK_i$  and  $SK_i$ ; denote the new certificate that user is currently applying for as  $Cert$ ; denote certificate issued to RA by  $CA_i$  as  $Cert_{RAi}$ , and corresponding public key and private key are denoted as  $PK_{RAi}$  and  $SK_{RAi}$ .

When a user owns a certificate  $Cert_i$  issued by  $CA_i$  belonging to  $G_{RA}$ , the application process of new certificate and RA's approval process are showed as following:

- 1) The user adds signature to the application information  $INFO$  with  $Cert_i$  on the client, that is  $SIGN(INFO)$ . Detailed steps are: compute the hash value of  $INFO$  to get the abstract  $INFO_{hash}$ , and encrypt  $INFO_{hash}$  with  $SK_i$ .
- 2) The client encrypts  $INFO$  with  $PK_{RAi}$ , the public key of RA, to get  $ENCRA(INFO)$ , and then join the signature and encrypted  $INFO$  together to get  $S$ :  $SIGN(INFO) \parallel ENCRA(INFO)$ . Finally client sends  $S$  to RA.
- 3) RA receives  $S$ , and resolves to get  $SIGN(INFO)$  and  $ENRA(INFO)$ .
- 4) RA verifies the signature using the following steps:
  - a) decrypt  $ENCRA(INFO)$  with  $SK_{RAi}$  to get  $INFO$ ;
  - b) decrypt  $SIGN(INFO)$  with  $PK_i$  to get  $INFO_{hash}$ ;
  - c) RA computes hash value of  $INFO$  and compares it with  $INFO_{hash}$ . If the two values are identical, RA implies the information is submitted by the user; otherwise, failure information is returned to user by RA.
- 5) If the verification of the signature above is successful, RA keeps on to extract user's personal information  $INFO'$  from the certificate  $Cert_i$ , and compare  $INFO'$  with  $INFO$ . If the key information items in both of them are identical, it manifests the submitted information is authentic and could be accepted by RA. Otherwise, RA returns signals to the user indicating the failure.

### 2.4 Features

Compared with traditional on-site application method, the approach proposed here can improve the checking efficiency of RA. It has the features as following:

1) Approval process is convenient. For users, online submission is convenient for local application; for RA, relying on certificates already owned by users, it can achieve the goal of automated verification and improve the efficiency of verification so that RA can significantly reduce the resource consuming.

2) A method of building horizontal chain of trust is proposed. Strictly hierarchical trust model of CA is vertical, and trust chain is therefore vertical. Inferior CA is authorized by superior CA[3]. However, a CA-Trust Set stands for horizontal trust chain, and may not set a trust anchor[4] of every CA in the set. There are no strict interoperability among them. This trusted set is application-oriented, and has more practical significance. This trust structure is more portable and practical.

3) Users can be hierarchically managed according to the categories of user information stored in Cert<sub>i</sub>. The reason, for that Cert<sub>i</sub> issued by CA in the CA-Trust Set can be used to check the authenticity of submitted key information when applying new certificate, is that there are some duplicated information items between information stored in Cert<sub>i</sub> and information submitted while applying for new certificate, such as user names and ID numbers and so on. The more Cert<sub>i</sub> that can testify the authenticity of submitted information there are, or the larger the union of user information of multiple Cert<sub>i</sub> is, the more authentic the submitted user information is. RA can hierarchically manage users by the trust degree of user information, and relevant digital certificates can also be assigned with corresponding level privilege.

### 3 Conclusion

This paper gives the definition of the CA-Trust Set, G<sub>RA</sub>. Based on the definition, a fast application and approval process of digital certificate is proposed. That is verifying the authenticity of submitted information through user-owned certificates issued by CA in the set of G<sub>RA</sub>. This approach enables user to apply for new certificates online and simplifies the application process. At the same time, it enables the server to check the authenticity of the submitted information automatically, so that the efficiency of information verification and the whole PKI system is significantly improved. In the future, research could focus on how to take further use of the existing certificates to build horizontal and specific application-oriented trust chain among different CA.

### References

1. Lu, J.: The Research and Design of Certification Authority Based on the Elliptic Curve Cryptography (2006)
2. Tong, L., Liu, L.: One of Arithmetic Models of Identity Trust Management Based on Dynamic Federation. Computer Technology and Development 20(2), 152–155 (2010)
3. Jiang, H., Cai, Z., Rong, X., Zhou, L.: Analysis and Research of Several Trust Models in PKI. Computer Measurement & Control 11(3), 201–204 (2003)

4. Steven, L.: PKI interoperability framework (February 20, 2010),  
<http://www.pkiforum.org/pdfs/PKIInteroperabilityFramework.pdf>
5. Zhang, F., Gu, Q., Jing, J.-W., Lin, J.-Q., Zha, D.-R.: PKI application interoperability evaluation. Journal of the Graduate School of the Chinese Academy of Sciences 27(6), 824–830 (2010)
6. Guan, Q., Zhua, Y.: The Research of PKI Trust Model and Interoperability. Computer Applications 26(9), 2148–2159 (2006)

# A Navigation System on the Dynamic Constraint Condition

Qiang Ma<sup>1</sup>, Jian Deng<sup>2</sup>, and Gan Wei<sup>3</sup>

<sup>1</sup> Qingyuan Polytechnic, Qingyuan, Guangdong, P.R. China  
martinstrong@163.com

<sup>2</sup> General Administration of Press and Publication of the P.R. China Beijing, P.R. China  
dj200601@hotmail.com

<sup>3</sup> Qingyuan Advanced Technical School, Qingyuan, Guangdong, P.R. China  
chweigan@hotmail.com

**Abstract.** For an intelligent navigation system, this paper focuses on creating an algorithm of the intelligent path computation, based on the dynamic constraint condition. Traditional navigation systems like to install GPS devices with vehicles. Normally the signals will not be sent back to traffic control center. The vehicles use the signals independently and locally. The dynamic constraint condition will copy the ant colony algorithm. It will study how to take advantage of all the GPS signals together, and how to solve the problem of path optimization.

**Keywords:** cloud computing, intelligent navigation, path computation, GPS (Global Position System).

## 1 Background of Traffic Control System

Dynamic traffic information service is always one of the hot issues for any intelligent navigation systems. In developed countries, governments and enterprises have done a lot to study the traffic information service, which is thought to be an effective solution to the traffic problems in metropolis. Among all the problems with an intelligent traffic system, three problems are always there. The first problem is how to collect the navigation data and send it back to a control center. The second one is the computation of mass data and the data mining for the traffic service. And the third one is path optimization based on the dynamic constraint condition.

Here we will mainly study the path optimization based on the dynamic constraint condition because the other two problems have been solved already. The first one has been solved due to the development of the mobile communication technology. The second one has also been solved by the technology of cloud computing, which is a new distributed computing model. Cloud computing providers can supply reliable and self-defined processing for mass data computation and data mining.

## 2 A Real Navigation Problem

### 2.1 Problem Description

It is usual that an electronic map cannot be updated in time and it cannot report the real-time traffic information.

The traditional GPS navigation system cannot report the real-time traffic information. Normally the GPS devices installed with vehicles receive synchronizing signals from satellites and compute the data for their coordination. The data and the electronic map together will guide vehicles to their destinations. However, with the growing number of cars and the changing of the road network, it is not so easy to get updated map that it is impossible to report the real-time traffic information.

Integrating the technology of mobile communication, the Internet of things, and cloud computing, the dynamic routing for navigation will achieve the target of intelligent traffic. Dynamic electronic maps and dynamic traffic information with terminals will be the key to the achievement. The following will discuss how to process dynamic traffic information.

## 2.2 System Architecture

Here are some descriptions for the concepts used in figure 1.

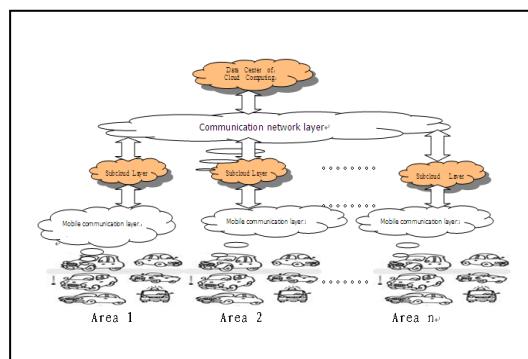
Data Center of Cloud Computing is used to schedule tasks among different types of clouds. It also administrates account information. The accounts here are authorized system users. They will provide the basic data for the intelligent navigation.

Network Communication Layer is used for the data communication between different cities. The security of data transmission should be assured in this level.

Subcloud Layer is to collect the information for moving vehicles within one area. The computation of mass data will be done here.

Mobile Communication Layer is used for the data transmission, by the mobile communication terminals, between vehicles and subcloud layer.

Application Layer is for dynamic map administration, the creation of dynamic traffic information with mobile terminals, the account operation, and the administration of mobile communication, etc.



**Fig. 1.** The above picture is about the system architecture of an intelligent navigation system, which is based on the dynamic constraint condition

## 3 The Effective Algorithm for the Traffic Problem

In any cases, a single vehicle can move only within an area. It means the data of many mobile terminals with the vehicles can report the web information of a road network.

By collecting and analyzing the data of GPS terminals in the road network, we can get the basic traffic information. By processing the basic data and then publishing the newest data, the dynamic traffic information can be sent to the mobile terminals and the electronic map can be updated.

The biggest issue here is the path optimization based on the dynamic constraint condition. To solve the problem we need two steps. The first one is the computation on the dynamic constraint condition. And the second one is the optimization of static weighted paths.

### 3.1 The Computation on the Dynamic Constraint Condition (for Server)

#### 3.1.1 The Theory about Ant Colony Algorithm

Let's see how ants move with their navigation. Ants that return from foraging journeys can use landmarks to find their way home, but in addition they have an internal backup system that allows them to create straight shortcuts back to the nest even when the outbound part of the forage run was broken or very winding. This backup system is called the 'path integrator' by computing the messages left by other ants.

The complex social behaviors of ants have been much studied by computer scientists. The problem of finding the shortest paths has become the field of ant colony optimization, which is one of the most successful and widely recognized algorithmic techniques based on ant behavior.

We can copy the ant's navigation system through the technology of telecommunication and computer networks.

#### 3.1.2 The Description of Ant Colony Algorithm

By the ant colony algorithm, we will show the solution of finding the shortest path for n cities in a map.

Assuming the number for ants are k ( $k=1,2,\dots, m$ ),  $P_{ij}^k(t)$  stands for the possibility of the ant k move from city i to j at time t.

$$P_{ij}^k(t) = \begin{cases} \frac{\tau_{ij}^\alpha(t)\eta_{ij}^\beta(t)}{\sum_{r \in allowed_k} \tau_{ir}^\alpha(t)\eta_{ir}^\beta(t)}, & j \in allowed_k \\ 0, & otherwise \end{cases}$$

Following is the variables in the formula.

$allowed_k$  is the group of cities before the ant k goes to the next city.

$\tau_{ij}(t)$  is the messaging consistency left by the ant k between city i and city j at time t.

$\eta_{ij}(t)$  is the enlightening message for the ant k move from city i to city j.

$\alpha$  stands for the importance of the messaging consistency.

$\beta$  stands for the importance of the enlightening message.

After n pieces of moment, ant k can go through all cities and the messages left for every path will like this.

$$\tau_{ij}(t+n) = (1 - \rho) \cdot \tau_{ij}(t) + \Delta\tau_{ij}$$

$$\text{and } \Delta\tau_{ij} = \sum_{k=1}^m \Delta\tau_{ij}^k.$$

$\rho$  is the factor of message left by ant k.

$\Delta\tau_{ij}^k$  is the messages volume left by ant k between city i and city j.

### 3.1.3 Steps of the Ant Colony Algorithm

Step1. Reset the iteration counter nc to zero,  $\tau_{ij}(0) = c$  (constant), and  $\Delta\tau_{ij} = 0$ .

All the ants will be in the starting point at the moment.

Step 2.  $P_{ij}^k$  is the possibility of ant k move from city I to city j. It will be input to a solution set.

Step 3. After n pieces of moment, ant k will finish all paths. Find the best solution in the set by computing every ant's path lengths.

Step 4. Renew the message consistency  $\tau_{ij}(t+n)$  for every path.

Step 5. Reset  $\Delta\tau_{ij}$  to zero, and reset nc with nc+1.

Step 6. If nc is less than the scheduled iteration frequency, then return to step2. Or else it will be the end of the steps. The best solution will be output and saved.

The ant colony algorithm is suitable to the distributed computing model. Comparing with the related algorithms such as genetic algorithm and simulate anneal arithmetic, it is still one of the best solution to the problem of the path optimization [1].

### 3.1.4 Improvement of the Ant Colony Algorithm

We need to improve the ant colony algorithm because the traditional algorithm does not consider some traffic conditions, such as the density of the vehicles during traffic jams/accidents.

An intelligent navigation system should be based on the dynamic constraint condition. We have to think about how to compute another shortcut when traffic jams/accidents happen, which are very common in our daily lives. For example, it happens frequently that all the vehicles have to stop in a smooth road, or in an intersection vehicles are not allow to turn right/left but have to go straight [2].

We can adopt Greenshield linear model to improve the ant algorithm. Greenshield model assumes that, under uninterrupted flow conditions, speed and density are linearly

related. It helps us to develop a model of uninterrupted traffic flow that predicts and explains the practical trends in real traffic flows.

Here is the model.

$$\begin{cases} v_{ij} = v_{ij}^f (1 - \rho_{ij} / \rho_{ij}^{jam}) \\ q_{ij} = \rho_{ij} v_{ij} \end{cases}$$

$v_{ij}$  stands for the speed of the vehicle at node i, and  $v_{ij}^f$  the free flow speed at node j.

$\rho_{ij}$  stands for the density of vehicles between node i and node j, and  $\rho_{ij}^{jam}$  the vehicle-blocking density between node i and node j.

$q_{ij}$  is the traffic flow magnitude between node i and node j.

While  $v_{ij}^m = v_{ij}^f / 2$  and  $\rho_{ij}^m = \rho_{ij}^{jam} / 2$ , the maximum traffic flow magnitude will be  $q_{ij}^{\max} = v_{ij}^f \rho_{ij}^m / 4$ . Among them,  $v_{ij}^m$  is the speed constant and  $\rho_{ij}^m$  the density constant.

If  $\rho_{ij} \geq k_{ij} \rho_{ij}^m$ , there is a traffic jam and the critical speed at the moment is  $v_{ij} \leq (1 - k_{ij}) v_{ij}^f$ .

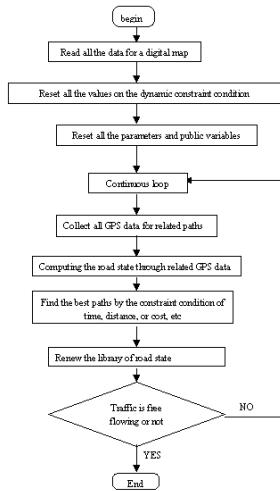
Considering the dynamic constraint conditions is an improvement to the traditional algorithm. When there are traffic jams/accidents, we cannot find the shortest path by only the ant colony algorithm, which will make the blocking more serious if vehicles are misguided to a blocked road.

### 3.2 The Optimization of Static Weighted Paths (for Client)

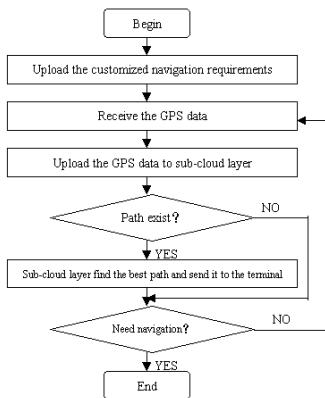
Clients need to be set up for a lot of functions, such as the automatic upgrade of electronic maps, the Automatic upgrade of the constraint condition, searching and switching between servers, autonomous navigation, and automatic navigation. Comparing with the traditional terminals, it seems there are too many added functions for the client terminals. However, servers will do most computing and there is no need for clients to take care of the computing task [3].

## 4 Algorithm Flowcharts

Please see the following figure2. We will see every vehicle as an ant. All vehicles in an area will like an ant colony. The algorithm flowchart is to track the movement of every vehicle, compute the road state, and then plan the driving path for vehicles [4]. The data process flowchart for client is shown in figure3.



**Fig. 2.** Algorithm flowchart for the path optimization



**Fig. 3.** Data process flowchart for client

## 5 Conclusion

Integrating the technology of mobile communication, the Internet of things, and cloud computing together, this paper discusses how to take advantage of using all the standalone GPS signals effectively. It also describes an algorithm on the dynamic constraint condition for an intelligent navigation system, which imitates ant's navigation system to solve the problem of the path optimization.

## References

1. Dorigo, M., Gambardella, L.M.: Ant Colony System: A Cooperative Learning Approach to the Traveling Salesman Problem. *IEEE Trans. on Evolutionary Computation* 1(1), 53–66 (1997)
2. Meuleau, N., Dorigo, M.: Ant colony optimization and stochastic gradient descent. *Artif. Life* (2002)
3. Dorigo, M., Gambardella, L.M.: Ant Colonies for the Traveling Salesman Problem. *IEEE Trans. on Evolutionary Computation, BioSystems* 43(2), 73–81 (1997)
4. Lee, S.G., Jung, T.U., Chung, T.: An effective dynamic weighted rule for ant colony system optimization. In: Proc. of the 2001 Congress on Evolutionary Computation, vol. 2. IEEE Press (2001)

# How to Use Oprofile to Improve an Algorithm for Skia in Android

Haihang Yu and Gang Du

Dept. of Information Engineering, China University of Geosciences, Beijing, China  
yhh5158@gmail.com, dugang@cugb.edu.cn

**Abstract.** With the development of Tablet PC, especially the rapid development of tablet PC based on android platform, more and more people started using the Tablet PC as an entertainment tool, while web browsing and reading text is a very important function of tablet PCs. Skia is a complete 2D graphic library for drawing Text, Geometries, and Images and is the engine of android graphic system. Most users like read in the case of the page resizing, so the optimization of skia library can let web browsing and text reading more fluently. The main purpose of this article is to measure the hot function through oprofile and optimize some certain ones, in the case of amplification of the page or text reading. Results show that the optimization can improve reading effectively.

**Keywords:** skia library, oprofile, Tablet PC.

## 1 Introduction

Android is a Linux-based, open source, smart phone operating system developed by Google, which includes operating systems, middleware and applications [1]. As the first complete, open, free mobile platform, Android is a hot topic in the industry since its introduction because of its excellent portability and powerful features as well as the good momentum in the application of embedded devices. Many Tablet PC and smart phones are based on the android, and in the aspect of graphics, use a 2D vector graphics library Skia specially modified for the Android.

## 2 The Introduction of Skia

The underlying of Android 2D system is achieved by the local library of skia and provides graphics interface to Application through the JNI. As a low-level graphics, images, animation, SVG, text and other aspects of the graphics of google, skia is the graphics system in Android engine. Skia is a C++ native code libraries. It contains three libraries: libcorecg.so, liblibsgl.so, libskiagl.so. Libcorecg.so is the most basic library of skia, which provides a set of basic features, such as some mathematical calculations, memory management and other infrastructure categories. It's also called by the other two library as a tool. liblibsgl.so is graphics library, which contains the graphics rendering, image codecs, effects and many other elements. libskiagl.so is a library associated with skia and openGL.

### 3 Introduce and How to Use the Oprofile

#### 3.1 The Introduction of the Oprofile

Oprofile is one of the several evaluation and performance monitoring tools used for Linux. It can work in different architectures, including IA32, IA64, AMD Athlon family, and ARM, etc. Oprofile is included in the Linux 2.5 and later kernels, also in most newer Linux distributions, which has been integrated in the Android[2].

With very low overhead, oprofile make performance analysis in function-level (function-level profiling)on all the running code (including kernel, kernelmodules, libraries, applications), track the call information whose function occupy CPU high, both to determine where there is need to optimize the performance bottleneck.

Oprofile supports two sampling mode: event-based sampling and time-based sampling. Event-based sampling is oprofile only record the occurrence of specific events (such as L2 cache miss), when it reaches the value the user set, oprofile record once (taken a sample). This approach requires an internal CPU performance counter. Time-based sampling is oprofile helped by the OS clock interrupt mechanism. During each clock interrupt, oprofile will record once (one sample taken), which is divided into RTC mode (RTCmode, for 2.2/2.4 kernels) and the timer interrupt mode (timer interrupt mode, for more than 2.6 kernel). The purpose of the introduction of sampling mode timer is providing support for the CPU with no performance counters. Its accuracy is lower than event-based sampling. Because it must rely on the support of OS clock interrupt, oprofile can not make analysis for disabled interrupt code.

In Android, oprofile is divided into two parts: target-side and host-side. Target-side runs on the device, including a kernel module (oprofile.ko) and a user-space daemon (oprofiled). The former is responsible for accessing the performance counters or registration the time-based sampling function (using register\_timer\_hook registered it, making the clock interrupt handling access it when it finally execut the profile\_tick), and then place core samples within the buffer. The later runs in the background, which is responsible for collecting data from kernel space, writing into the file. Host-side runs on a PC, including a post-processing tools for generating readable analysis from the original sample data.

#### 3.2 How to Use It

When you use the oprofile, first, you should have the root permissions and second the space of data partition is enough. Then you can do as follows.

- Set up the phone target

Push the oprofile.ko to the data partition:adb push oprofile.ko /data.

- Calculate kernel virtual address range

If you want to oprofile the kernel,you should execute this step, whether you can skip this step.

The value of the \_text is the starting address of the kernel and the value of the \_etext is the ending of the kernel which you can find in the file of kallsyms.

- keep the cpu frequency performance

You should keep the cpu performance before you start to sample to make sure the results is correct and consistency.

echo performance > /sys/devices/system/cpu/cpu0/cpu-freq/scaling\_governor.

- Configurate the oprofile

You should configurate the directory of the oprfile and vmlinux,the range of kernel etc.  
insmod /data/oprofile.ko opcontrol --setup oprofiled --session-dir=/data/oprofile  
--vmlinux=/data/vmlinux--kernel-range=**start,end**

--events=CPU\_CYCLES:255:0:50000:0:1:1 --separate-lib=1 --separate-kernel=1(the **start** is the starting kernel address and the **end** is the ending).

- Begin samplimg

You can use opcontrol --start to begin smapling ; you can also use opcontrol --status to know how many have you sampled.

- Stop sampling

When the sample size is more than 3000,you can stop sampling use opcontrol --stop.

- Upload the data and generate results

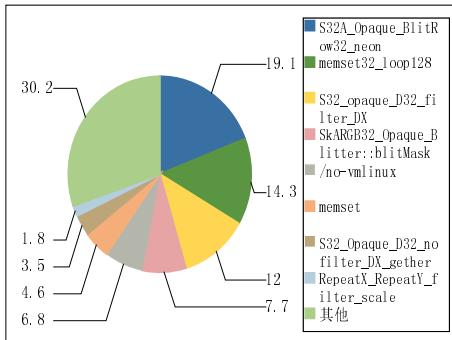
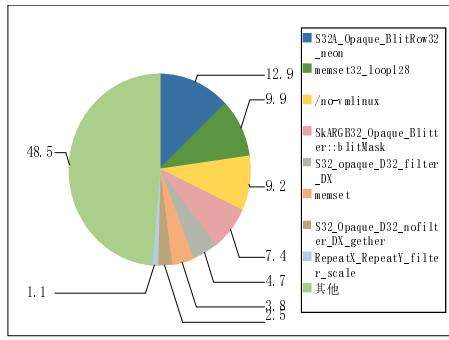
You can use opimport\_pull to pull the data from the pad and then use oreport to generate the results.

## 4 The Smage Resizing in the Original Skia Library and the Result of Oprofile

The draw of all the 2D images including the image resizing will call the skia library in the android. There are many images in the paper whether to browse the web or reading a text, and partly the paper resizing is the image resizing. Image resizing is the zoom in or zoom out of the image and is widely used in the image processing technology. Image magnification(for up sampling) is the improvement of the resolution ;while the narrowing of the image(for down sampling) is the reduction of the resolution. The resolution of the images is the number of the dotmatrix which is united as dip[3]. The common resolution includes 640 \* 480、1024 \* 768、1600 \*1200、2048 \* 1536. The former is the width ,the latter is the height, and the pixels of the image is multiplied by both[4].The scaling principle of image is based on the original pixels of the image, in the constraint condition of retaining the original basic image information as much as possible, processing the existing pixels number according to certain rules of operation to achieve the goal that the number of pixels are increased or decreased so as to realize the process of amplification and reduction of the image[5].

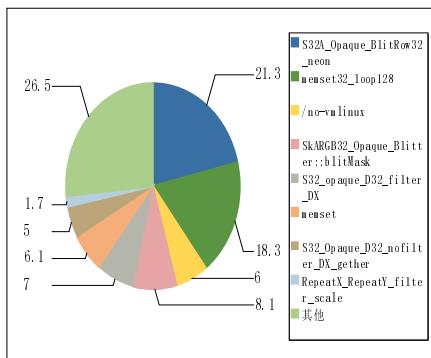
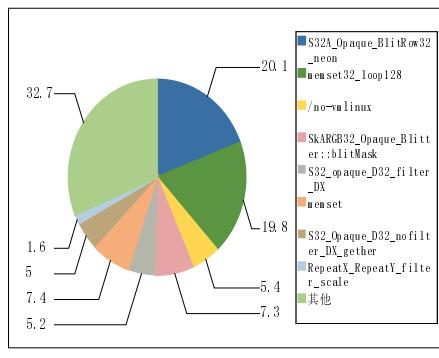
There will be two processes form the original image to the final image which is resized in the skia. First, the calculation of matrix from the original to the destination which is the function of the mproc (state, buffer, n, x, y);Second, the real realization which is the function of sproc (state, buffer, n, dstC). In the above two functions the state is a object of the SKBitmapProcState class, the buffer is an array, n is the pixels, x、y is the coordinate of the destination, dstC is the address of the destination pixel which includes the value of the three basic color (for R、G、B). You can also download the source code of skia to realize the process of the function [6].

In enlarge sina homepage by draging down the page get profile analysis structure as shown in figure 1, figure 2.

**Fig. 1.** The result before optimization(1)**Fig. 2.** The result before optimization(2)

## 5 Optimization and the Oprofile Results after Optimization

From the aboving pie chart and the analysis of the source code, we can see that the two mentioned functions, the mproc function is corresponding to RepeatX\_RepeatY\_filter\_scale, However, the sproc function is corresponding to S32\_opaque\_D32\_filter\_DX. We can also get the analisis that under the Enlarge page the calculations of transition matrix don't have the biggest impact but the realization of the conversion process. From the view of optimization, optimizing sproc will have more distinct results than optimizing mproc, so we should pay more attention to optimizing sproc.

**Fig. 3.** The result after optimization(1)**Fig. 4.** The result after optimization(2)

From skia's source code we kown that each executive of function sproc (state, buffer, n, dstC), are all take n pixel operations. And by printing time stamp in the source code, it takes up most the CPU resources in the whole process. Can we try to reduce the operation? This is the major breakthrough to optimize. Generally when we browsing the webpages, they will download many pictures .And those images on a rectangular area the pixel color are the same, so Don't need operate n pixels but only take 1 pixel

operations. If the values of the three base colors are same in a pixel, we can say the picture color of those n pixels in the bitmap is the same. So do not have to go to operations in addition n-1 pixel. using `sk_memset32()` function directly provided skia bank, Realizing the replication from the original figure to the target one. If the pixel value in one elements of operation are not equal, then operating according to the original algorithm. After optimization, in enlarge sina homepage by draging down the page get profile analysis structure as shown in figure 3, figure 4 .

## 6 The Increased Performance Analysis after Optimization

If only see the proportion of `S32_opaque_D32_filter_DX`, and don't see the actual effect of optimization, this is because when you drag, drag frequency will affect oprofile analysis results. So we can't just look at this a ratio, and sould find a starting point. And function `SkARGB32_Opaque_Blitter::blitMask` don't going optimization, so we can take it as a benchmark to analyse the improvement of the performance of the optimization. In the original two sets of data, the ratio of the proportion of `S32A_Opaque_BlitRow32_neon` and the proportion of `SkARGB32_Opaque_Blitter:: blitMask` respectively is  $12:7.7 = 1.56$ ,  $7.4: 4.8 = 1.54$ , the ratio A1 of the proportion of `S32A_Opaque_BlitRow32_neon` and the proportion of `SkARGB32_Opaque_Blitter:: blitMask` approximately equal to 1.55. But after optimization the ratio of the proportion of `S32A_Opaque_BlitRow32_neon` and the proportion of `SkARGB32_Opaque_Blitter:: blitMask` in two groups of data respectively is  $7:8 = 0.87$ ,  $5.2:7.3 = 0.71$ , the ratio B1 of the proportion of `S32A_Opaque_BlitRow32_neon` and the proportion of `SkARGB32_Opaque_Blitter:: blitMask` approximately equal to 0.8. Comprehensive promotion effect  $(A1-B1)/B1$  is approximately equal to 1.

## 7 Conclusion

This paper present the improvement algorithm for skia is effective. After the improvement the proportion of the sproc was reduced evidently and the page drag effect improved strongly. It could be summarized as follows. The skia as a 2D graphic library for drawing Text, Geometries, and Images, if the optimization of skia is well done, the user experience also will be improved well.

**Acknowledgment.** Thanks for the support of the Fundamental Research Funds for the Central Universities, the number is 2-9-2011-229. The author gratefully acknowledges associate professor DU Gang for his helpful suggestions and comments.

## References

1. Innovaspire, Android:Change the mobile landspace. In: IEEE Pervasive Computing, vol. 10, pp. 4–7 (January–March 2011)
2. Roberta, G., Fabrizio, P., Kei, D.: Analysis of system overhead on parallel computers. In: Fourth IEEE International Symposium on Signal Processing and Information Technology, ISSPIT 2004, December 18–December 21. Institute of Electrical and Electronics Engineers Inc. (2004)

3. Yin, Z., et al.: Fractal interpolation image enlarging and compression. In: SP 1998, vol. 14(1), pp. 86–89 (1998)
4. Atkins, C.B., Bouman, C.A., Allebach, J.P.: Tree—based resolution synthesis. In: Proceedings of IEEE ICIE 1999, pp. 405–410 (1999)
5. Avidan, S., Shamir, A.: Seam Carving for Content-Aware Image Resizing. ACM Transactions on Graphics 26(3) (2007)
6. Christian, H.: Generic method for 2D image resizing with non-separable filters. In: Proceedings - International Conference on Image Processing, ICIP, vol. 3, pp. 1653–1656 (2004)
7. Shih, G., Lakhani, P., Nagy, P.: Is Android or iPhone the Platform for Innovation in Imaging Informatics. Journal of Digital Imaging 23, 2–7 (2010)

# Cloud-Based Service Composition Architecture for Internet of Things

Li Liu<sup>1</sup>, Xinrui Liu<sup>2</sup>, and Xinyu Li<sup>3</sup>

<sup>1</sup> School of Automation & Electrical Engineering, University of Science& Technology Beijing,  
100083 Beijing, China

<sup>2</sup> School of Economics & Management, University of Science and Technology Beijing,  
100083 Beijing, China

<sup>3</sup> School of Computer & Communication Engineering, University of Science and Technology  
Beijing, 100083 Beijing, China  
Liuli@ustb.edu.cn, {ustb\_lxr,Lxyu77}@163.com

**Abstract.** IoT is heterogeneous, multi-layer, distributed network, composed of a variety of network interconnected. In such infrastructures, composed of a large number of resource-limited devices, the discovery of services and on demand provisioning is a challenge. The emergence of new IoT services will further promote the development of service computing related disciplines. In this paper, cloud based service architecture for IoT is proposed and a several key technologies are explored, such as light-weight semantic of service, context-aware based service discovery mechanisms and adaptive service composition model. We hope that our proposals will provide some key inputs for realization of IoT.

**Keywords:** cloud-based architecture, internet of things, service composition model.

## 1 Introduction

With the development in information and communication technology, communication networks as an important foundation of information and communication technology, has moved from person to person communication development to communication between people and objects, and things to things communication, moreover gradually moving from the vertical local things and things connected to the horizontal cross-application, cross-boundary Internet of Things (IoT)[1].

In recent times, the IoT has developed rapidly and globally due to increasing projects invested by government and enterprise. IoT are considered as more information on the Internet to the extension of the physical world through a variety of sensing, detection, identification, location tracking and monitoring equipment [2]. This is just the first stage of human society to achieve "sensing, aware, control" for the physical world. Internet-based computing intelligence services and the feedback and control of physical world are other two important aspects.

Computing technology is developed to large-scale, high-performance and distributed cloud computing on the one hand, and through cloud platform to provide

users with "on-demand rental" of IT services. On the other to the development of ubiquitous computing, a variety of embedded smart devices through wireless network to provide "any time and anywhere" ubiquitous service terminal for cloud computing. With the development of "cloud" as Internet information infrastructure, service computing, is more important.

The optimization strategy and intelligence services adapt to IoT application will be more in the form of service composition. Services composition has carried out extensive research on a variety of aspects, such as domain ontology [3], objective assessment, semantic-based [4] and Quality of Service (QoS) oriented service composition [5].

Real-time sensors, highly concurrent, independent collaboration and other features of IoT raise new challenges for service computing. In such wireless mobile environment, network bandwidth, memory capacity, processor, power and other resources of embedded devices are very limited. The existing Web service-oriented architectures (SOA) are designed for PC class devices, there is a lack of light-weight approaches suitable for the resource constraints of embedded devices. So, embedded Web Services, mobile wireless Web services, and service-oriented mobile devices are researched [6,]. Some work start from the service component, focused on SOAP protocol for low resource consumption, such as gSOAP and kSOAP.[7] Currently, the embedded Web services is still in development stage, the academic community do a lot of exploration in the service component, middleware, development tools, and so on [8]. More promote works are done by industry. For example, Microsoft strengthens the support services with limited resources in Windows Embedded CE 6.0 R2 based on embedded Web services. Specifications, include OSGi Service Platform Specification (OSGi), J2ME Web Services Specification(JSRI72), and Devices Profile for Web Services (DPWS). These standards have adopted a hierarchical structure to support equipment service and interconnected. Additionally, real-world services are found in highly dynamic environments where devices and their underlying services degrade, vanish, and possibly re-appear. This implies the need for automated, immediate discovery of devices and services as well as their dynamic management.

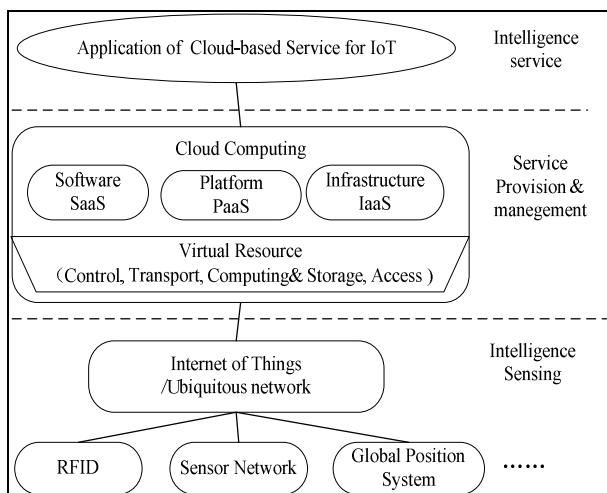
Due to characteristics of IoT and the limitations of traditional service composition, we propose the cloud-based service architecture for IoT and explore the key technologies within the framework.

## 2 Cloud Based Architecture for IoT

Cloud computing is an innovative mode for IT resource management and computing method, which can manage a shared resource efficiently and flexibly in a self-service way. Cloud computing technology promotes availability of resources; self-service provisioning on-demand and quantitative service are essential features. Cloud computing technology can provide a unified service delivery platform for the IoT applications. More devices in IoT will be connected to a shared resource pool, attaching to the cloud for storing and retrieving of data. The cloud-based architecture

of IoT efficiently support various smart services. We name the cloud-based IoT service environment as “Cloud-based Service of IoT”. These advantages for the proposed service pattern might come from characteristics of cloud computing depending on specific use cases of the IoT, such as: flexibility of resource allocation, scalability, virtualization, ubiquitous, etc.

Cloud computing based cross-layer architecture for IoT need to be studied by means of expanding the existed architecture of IoT to increase ubiquitous service for cloud computing. A basic mechanism combining the interaction of physical movement and mobile services need to be proposed to improve the interoperability. In the cloud based service architecture for IoT, a novel resource management for service provisioning and management will be a key enabler for realizing smart services of the IoT. In the IoT environment, the memory capacity of equipment, computing power of processor, power and other resources are so limited that Cloud-based Services of IoT are deployed in resource constraint equipments. The specification in the lightweight service-oriented semantic description is also to be researched, achieving the minimum service cost in the case of limited resources. Cloud-based service architecture for IoT is shown in Fig. 1.



**Fig. 1.** Cloud-based Service Architecture of IoT

In the environments of combining the cloud and the IoT, various intelligent services can be supported. Such a web-based service environment will allow not only creation of community-type services but also building of an open service platform environment which have features of interactive, collaborative and customizable on demand. There are two solutions for virtual resources of service provisioning and management in a Cloud-based Service of IoT environment: horizontally with network virtualization and vertically with resource virtualization. For network virtualization, it is essential to develop the technology that enables the creation of logically isolated network partitions over shared physical network infrastructures so that multiple

heterogeneous virtual networks can simultaneously coexist over the shared infrastructures. For resource virtualization, we need to consider the virtualization of resources including software, equipment, platform, computing, storage, etc.

### 3 Key Technology of IoT Service Composition

#### 3.1 Light-Weight Semantic for IoT Service

The semantic technologies are efficiently bridging information between global, enterprise, and embedded systems, use of ontology for cross-domain systems' organization and for interoperability in heterogeneous environments, dynamic reconfiguration capabilities, and adaptive resource management. The Internet of Things should become in fact the Semantic Web of Things.

IoT service is deployed in resource-limited terminal device with limited computing, communications and storage capacity. The existing service-oriented architecture is designed for the PC class devices and lack of lightweight solutions for resource-constrained equipments, such as embedded devices, so unified lightweight semantic descriptions for cloud services of IoT need to be researched. We can also study the semantic specifications associated with physical sensing layer, network layer, service structure layer, intelligence service application layer in the Cloud-based Service of IoT environment. Meanwhile, we can study an approach to describe dynamics of cloud services of IoT, focus on solving the ambiguity and uncertainty of the service description and form a semantic-based service description framework.

Due to IoT terminal devices with limited memory capacity, computing power of processor, power and other limited resources, the semantic description of lightweight services need to be studied to answer how to create a cloud service environment in these limited resources of RFID or embedded devices. OSGi specification supports service management with limited resources strongly and provides perfect security mechanism, service discovery mechanism. OSGi specification is formulated by OSGi Alliance which is an open standards organization, aiming at developing a set of open standards for embedded equipment, small mobile devices to provide seamless connectivity standard and achieve managed services. We can strengthen the capability of ontology-based service semantic description in the OSGi framework, focus on solving the ambiguity and uncertainty of the service description. Using the relationship between the ontology concepts to express the corresponding service semantics can enhance the ability to describe service with limited resources. We can establish a lightweight semantic descriptions system of IoT cloud service based on OSGi service specifications.

#### 3.2 Service Discovery Mechanism Base on Context-Awareness

Sensor networks are dynamic in the Cloud-based Service of IoT environment, which services provided by nodes are influenced by the changes of location, network and power. This change makes services dynamically applied to mobile nodes and new environmental changes. IoT services are in highly dynamical environment, services

may suddenly disappear or appear at any time, and the service capacity is constantly changing, automatically and immediately discover services are needed, so as dynamical management. We need to study dynamic discovery of services in the cloud environment, taking into account of context information, establish and optimize the model of service-time-effect. By tracking the services process, implement hot deployment of services and dynamic switching are the key to reflect dynamic adaptive capability of service discovery.

In the Cloud-based Service of IoT environment, the services need to continuously adapt to the changing situation, we intended to identify and locate context awareness services based on cognition and semantic interaction, monitoring services' implementation and migration. Context information directly influences services discovery, such as the status of equipment and services, service and the user's geographical location, the time-effect and the power of the node. We should take into account of context information to establish and optimize the model of service-time-effect. Using semantic ontology, fuzzy queries based powerful reasoning engine and smart tools extend and enhance currently existing service discovery protocol, applying semantic matching algorithms to select neighbor nodes. Implement services discovery based on mobile agent (MA), MA works as service spider, freely roam throughout the IoT environment, find web services and collaborative call these services to accomplish users' tasks.

### 3.3 Service Composition Model

We should research description of service component in the cloud environment, establish a targeted service combined mathematical model and study its various methods of refinement and optimization analysis on typical model. For the uncertainty of service composition, use stochastic process algebra to analyze both in qualitative and quantitative ways, giving strategy of improving system uncertainty. Further study assessment and monitoring of QoS measurement technology of the service composition to implement a service provision on demand.

The Cloud-based Service of IoT is heterogeneous, multi-level, distributed network consisting of multiple interconnected network. The services composition model is complicated, state space explosion problem is difficult to solve complex mathematical model, so research on equivalent simplification method of service composition model is key to whether service composition model analysis can be practical. Furthermore, stochastic theory or model such as stochastic process, stochastic high-level Petri net and stochastic process algebra can be used to establish evaluation model for QoS oriented service composition in the Cloud-based Service of IoT environment.

## 4 Conclusion

The features of IoT, such as real-time sensing, a high degree of concurrency, make the services collaboration in IoT increase a new challenge to service computing. We base on cloud computing technology to provide a unified applications service platform for IoT, which more embedded device can be connected to a shared resource pool, and

storing and retrieving of data to effectively support more types of intelligence services. The cloud-based IoT service architecture is proposed which combines both the cloud computing and the IoT. Furthermore, we present key points in this architecture, such as Light-weight semantics of services, service discovery mechanism base on context awareness, and service composition model. The proposed cloud-based IoT architecture aims to efficiently support varies services using cloud technology from different kinds of objects.

**Acknowledgments.** This work is supported by the National Natural Science Foundation of China under grant No.60873193 and No. 61001110.

## References

- [1] International Telecommunication Union UIT. ITU Internet Reports 2005: The Internet of Things (2005)
- [2] Chaves, L.W.F., Zoltán, N.: Breakthrough Towards the Internet of Things. In: Ranasinghe, D.C., et al. (eds.) Unique Radio Innovation for the 21st Century. LNCS, vol. 1, pp. 25–38. Springer, Heidelberg (2010)
- [3] Xiong, J., Hu, S., Liu, H.: On-Demand Service in Cloud Computing. ZTE Communications 16, 13–17 (2010)
- [4] Qiu, T., Hu, X., Li, P.: A Semantic Matchmaking System Mechanism for Web Service Discovery Based on OWL-S. Acta Electronic Sinica 38, 42–48 (2010)
- [5] Mokhtar, S.B., Georgantas, N., Issarny, V.: COCOA: Conversation-based Service COnnection in Perv Asive Computing Environments with QoS Support. Journal of Systems and Software 80, 1941–1955 (2007)
- [6] Guinard, D., Trifa, V.: Towards the Web of Things: Web Mashups for Embedded Devices. In: Proc. Workshop Mashups, Enterprise Mashups and Lightweight Composition on the Web, pp. 1–8. ACM, Spain (2009)
- [7] Terguieff, R., Haaajanen, J., Leppinen, J., Toivonen, S.: Mobile SOA: Service orientation on lightweight mobile devices. In: Proceedings of IEEE International Conference on Web Services, pp. 1224–1225. IEEE Computer Society Press, New York (2007)
- [8] Dominique, G., Stamatis, K., Domnic, S.: Interacting with the SOA-Based Internet of Things: Discovery, Query, Selection, and On-Demand Provisioning of Web Services. Transactions on Services Computing 3, 223–237 (2010)

# Internet of Things Applications in Bulk Shipping Logistics: Problems and Potential Solutions

Xin Song<sup>1</sup>, Lei Huang<sup>1</sup>, and Stefan Fenz<sup>2</sup>

<sup>1</sup> School of Economics and Management, Beijing Jiaotong University, Beijing, China

<sup>2</sup> Institute of Software Technology & Interactive Systems,

Vienna University of Technology, Vienna, Austria

a0430666@tom.com, stefan.fenz@tuwien.ac.at

**Abstract.** Internet of Things (IoT) technology can be used to significantly increase the efficiency of bulk shipping logistics operations. In this paper we provide an overview of common IoT shipping logistics applications, outline their main problems (lack of standards, costs, and security), and provide potential solutions to these problems. The research results support researchers as well as practitioners at designing and deploying efficient bulk shipping logistics systems.

**Keywords:** Internet of Things, shipping logistics, RFID, security.

## 1 Introduction

The port is an important hub of modern logistics industry, and it is the intersection point of the sea transport and land transport. With China's continuous economic development, the transportation capacity of energy, raw materials and other basic goods is still increasing. The growing throughput requires increased water transportation infrastructure, and calls for the informatization of bulk shipping logistics.

The Internet of Things (IoT), which is called the world's third wave of development in the information industry following the computer and the Internet, has been introduced to the waterway transportation in the container business, in order to improve the speed and accuracy of business information transfer. However, several problems at IoT applications in bulk shipping logistics exist. In this paper, we (i) briefly introduce the IoT, (ii) provide an overview of common IoT applications in shipping logistics, (iii) outline their main problems, and (iv) provide potential solution concepts to these problems. The identified solution concepts support bulk shipping logistics in:

1) Improving the information transfer efficiency between loading and unloading port.

2) Achieving comprehensive monitoring of freight during transportation. Achieving date exchange automatically between the ports and other shipping logistics companies, such as the railway company, the port authority, and the Customs.

3) Achieving automated loading and unloading in the terminal.

## 2 The Internet of Things

The Internet of Things (IoT) is a novel paradigm that is rapidly gaining ground in the scenario of modern wireless telecommunications. The basic idea of this concept is the pervasive presence of a variety of things or objects – such as Radio-Frequency identification (RFID) tags, sensors, actuators, mobile phones, etc. – which, through unique addressing schemes, are able to interact with each other and cooperate with their neighbors to reach common goals [1].

The IoT is a new vision of the future technological ubiquity in ubiquitous computing and communication era radically transforming the society, companies, communities, and personal spheres.

The IoT has the following three characteristics. The first characteristic is the comprehensive perception which is implemented by RFID, sensors or two dimensional barcodes. The second characteristic is the credible and real-time transfer. The information of things is accurately transmitted by various telecom networks and the Internet.

The last characteristic is the intelligent processing. The IoT uses various kinds of intelligent computing technology, such as cloud computing and fuzzy recognition to manage and analyze large pools of data and information.

From a technical perspective, IoT can be divided into three layers: the perception layer, the network layer and the application layer [2].

Perception layer: acquiring and perceiving the real data of the physical world, including identification, physical quantity, sound, images, which mainly involves RFID, sensors, bar codes and other technologies.

Network layer: through sensor networks, mobile communications and the Internet, the information from the perception layer can be transmitted accessibly, reliably and safely.

Application layer: the data from the network layer can be processed base on cloud computing. It can be used in intelligent identification, locating, tracking, monitoring and management.

In the following section we provide an overview of the most common IoT applications in shipping logistics.

## 3 Applications of IoT in Shipping Logistics

Through our previous research in the field of bulk shipping logistics, we conclude that the IoT technology can be used in the following three fields [3-5]:

### 3.1 Improving the Information Transfer Efficiency between Loading and Unloading Port

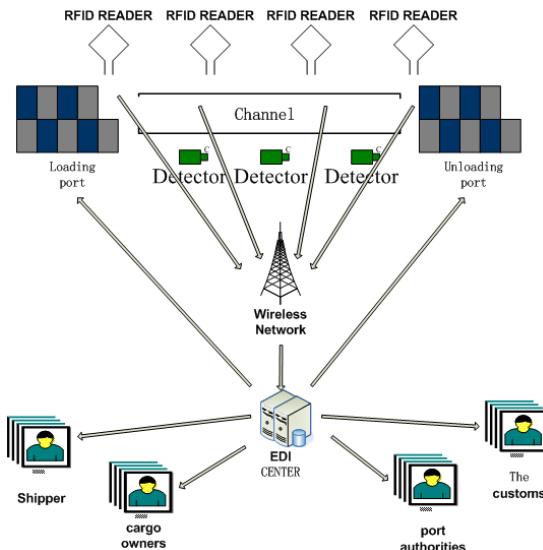
Because of missing standardized data exchange interfaces, ports have to use paper documents to exchange freight information. This practice cannot guarantee real-time data transfer and is prone to transmission and translation errors.

The problem can be addressed by using IoT technology and standardized data exchange standards. Based on a data exchange standard that is set up by the ports or standardization organizations, an Electronic Data Interchange (EDI) Center has to be set up to process and store the cargo information.

In an operative setting the loading port assigns RFID tags to the bulk freight. The unloading port operates RFID readers to automatically extract freight information from the bulk freight's RFID tags. The data is transferred in real-time to the EDI Center for further processing. Such a system removes the need for manually processing paper freight documents and improves the overall efficiency of loading and unloading processes.

### 3.2 Achieving Comprehensive Monitoring of Freight During Transportation

Using RFID tags not only solves the data exchange problem, but also can provide an effective way to achieve comprehensive monitoring of freight during transportation. Through the RFID readers which are installed next to the channel, the information of RFID tags can be read and sent to the EDI center with location information of the channel position. In addition, weather and water information can also be perceived by special detectors and send to the EDI center. Figure 1 shows how IoT technology can be used to track freight during transportation.



**Fig. 1.** IoT framework for tracking freight

The EDI center is the core of the framework, which is in charge of collecting cargo, channel and weather information and sending it to the special waterway transportation departments, such as port authority and the customs. By using this framework relevant stakeholders are enabled to retrieve real-time freight tracking information.

### 3.3 Achieving Automated Loading and Unloading in the Terminal

Figure 2 shows how IoT technology can be used to achieve automated loading and unloading in the terminal. This system depicted in Figure 3 includes the RFID tag and reader, an infrared detector and a license plate recognition system. When the delivery vehicle is on the weighbridge, the RFID reader, camera and infrared detector are

automatically opened, and the weighbridge begins to weigh the cargo. The vehicle weight, license number and RFID information are perceive and stored into the port management system automatically. After weighing, drivers can print the weighing ticket themselves and get off the weighbridge. If the management system finds problems during the weighing process, system will point the problem on the display. So the terminal can achieve automatically loading and unloading process, and realize unattended weighbridge.

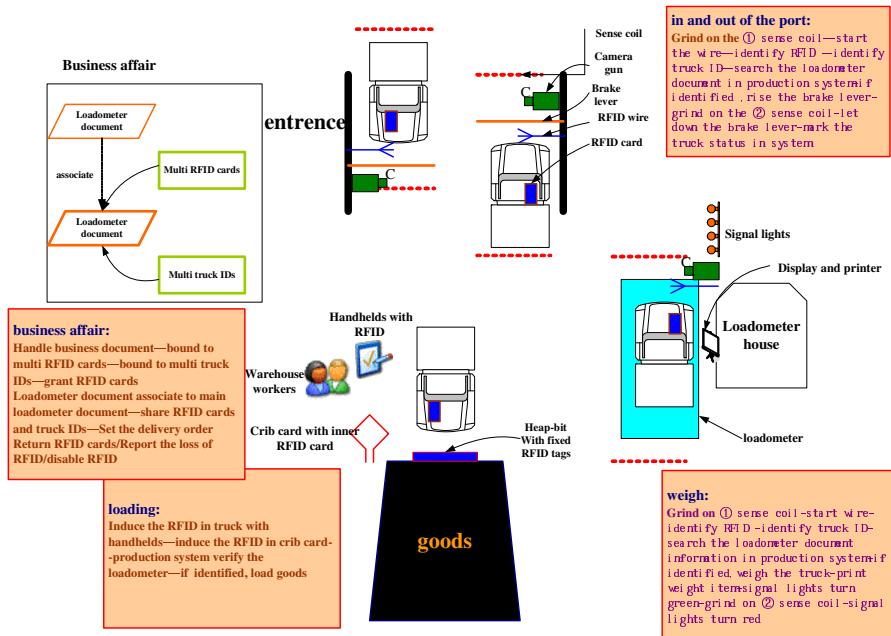


Fig. 2. Truck delivery process supported by IoT

## 4 Main Problems of IoT in Shipping Logistics and Potential Solutions

Nowadays IoT technology is used in bulk shipping logistics, in order to strengthen supervision and improve efficiency. However because of the diversity of the shape, packing, loading and unloading methods of bulk, there is still no practical application of IoT technology in bulk shipping logistics. In our research we identified three main problems at using IoT technology in bulk shipping logistics.

### 4.1 Lack of Standards

While the EDI standard is quite mature there exists no specialized data exchange standard for bulk shipping logistics. As the Bill of Landing (B/L) is the most important document in bulk shipping logistics we researched on a B/L data exchange standard. Table 1 shows our proposal for a structured B/L.

**Table 1.** Proposed B/L data structure

Field Name	Remark	Field Name	Remark
ID	the primary key	CARGOVOLUME	cargo volume
RFIDNUM	the RFID number	CARGOFLOW	cargo flow
BLFORMID	B/L number	DANGEROUSCARGO	is dangerous or not
SHIPUNIQUECODE	unique code of ship	DECLARATIONNUM	customs number
CONSIGNEE	consignee	CARGOLENGTH	cargo length
CONSIGNER	consigner	CARGOHEIGHT	cargo height
CARGOID	cargo tape code	CARGOWIDTH	cargo width
MARK	cargo specification	TRADEPATTERNS	trade type
PACKAGETYPE	package type code	ACTIVITY	business Type
AMT	cargo amt	FROMPORT	loading port
PIECEWEIGHT	average weight	TOPORT	unloading port
WEIGHTVALUE	total weight	VOYAGE	voyage
WEIGHTUNIT	cargo unit	ISAVAILABLE	the RFID number is available or not

## 4.2 Costs

The costs of IoT application in bulk shipping logistics include three aspects: (i) RFID tags, (ii) sensing equipment, such as RFID reader, infrared detector, and (iii) informatization equipment, such as servers, computers. The cost of the second and the third part are relatively fixed, so the cost reduction of RFID tags has the most significant impact on the overall IoT costs. Based on our research results we found two methods to reduce RFID tag costs:

- Instead of marking every single good with an RFID tag, we propose to use only one RFID tag per Bill of Landing to significantly reduce RFID tag costs in waterway transportation.
- Reuse of RFID tags: in our proposed B/L data structure (see Table 1) we included the field 'ISAVAILABLE'. When the freight is loaded to the ship at the loading port, the field is set to 'N'. When the ship arrives at the unloading port, the RFID reader beside the berth reads the RFID tag's information and sends it to the EDI center. If the unloading port is equal to the value of 'TOPORT', the value of 'ISAVAILABLE' is changed to the 'Y' and the RFID tag can be reused.

## 4.3 Security

By using RFID tags to enhance the efficiency of bulk shipping logistics operations we are faced with the following security problems [6] [7]:

- Corporate espionage threat: using unsecured RFID infrastructure makes it easier for competitors to remotely gather supply chain data.

- Infrastructure threat: the logistics processes highly depend on easily jammed radio frequency signals.
- Threat perimeter threat: large volumes of logistics data are electronically available outside the company boundaries and are therefore highly exposed to data theft or manipulation.
- Location threat: tags on valuable goods can be monitored and attackers can easily reveal their location.

The following basic security measures can be used to address the identified problems:

- Encryption of stored and transferred data. Potential problems include key management and increased tag costs if the encryption is directly performed on the RFID tag.
- Mutual authentication between RFID tag and reader. With authentication mechanisms in place, RFID tags only respond to authenticated readers and vice versa. Efficient key management solutions are crucial to enable mutual authentication.
- Tag pseudonyms. Whenever a tag is scanned it returns a different pseudonym and therefore makes it hard for an attacker to track specific RFID tags. At the system's backend the pseudonyms are mapped to the actual RFID tag ID.

By implementing encryption and authentication schemes in bulk shipping logistics systems we can prevent most of the identified threats. However, the risk of simply jamming the RFID infrastructure by radio frequency signals remains. Implementing efficient key management solutions which enable encryption and authentication remains the biggest challenge at implementing secure IoT bulk shipping logistics systems.

## 5 Conclusion

Internet of Things technology has a high potential of increasing efficiency in bulk shipping logistics. In this paper we identified the main IoT application areas in bulk shipping logistics and outlined the main problems: (i) lack of standards, (ii) costs, and (iii) security. Based on our research results we provided potential solutions for each of the problems: (i) a B/L data structure to move from paper-based Bills of Landing to electronic versions, (ii) cost reduction by decreasing the necessary amount of RFID tags and by reusing RFID tags after unloading operations, and (iii) basic security strategies to secure transmitted and stored shipping logistics data. In further research we aim at evaluating and refining our concepts in real-world scenarios with our partner companies.

**Acknowledgment.** The authors wish to acknowledge the support of the project of Guangdong Province Education Ministry: product business system of Guangzhou Port Group and general software industry (Project No. 2008B090500244) and the project of RFID-based vehicle management system application demonstration project (Project No. 2009B090300467).

## References

1. Sarma, A.C., Girao, J.: Identities in the Future Internet of Things. *Wireless Personal Communications* 49(3), 353–363 (2009)
2. Song, X., Huang, L., Wang, Q.: The Research of Internet of Things in Railway Transportation. In: Asia-Pacific Conference on Information Network and Digital Content Security, pp. 369–373 (2010)
3. Yan, B., Huang, G.W.: Application of RFID and Internet of Things in Monitoring and Anti-counterfeiting for Products. In: International Seminar on Business and Information Management (ISBIM 2008) (December 2008)
4. Huang, P.: RFID Technology in the Port Logistics Application. *Logistics Engineering and Management* (11) (2009)
5. Li, D.: RFID Technology in the Port Informatization Construction Application. *China New Technologies and Products* (20) (2009)
6. Juels, A.: RFID Security and Privacy: a Research Survey. *IEEE Journal on Selected Areas in Communications* (24) (2006)
7. Garfinkel, S.L., Juels, A., Pappu, R.: RFID Privacy: an Overview of Problems and Proposed Solutions. *IEEE Security & Privacy* (3) (2005)

# **Design and Research of Urban Intelligent Transportation System Based on the Internet of Things<sup>\*</sup>**

Hong Zhou<sup>1</sup>, Bingwu Liu<sup>1</sup>, and Donghan Wang<sup>2</sup>

<sup>1</sup> School of Information Science & Technology, Beijing Wuzi University, Beijing, China

<sup>2</sup> School of Media & Management, Communication University of China, Beijing, China

David\_csharp@yahoo.com.cn, Liubingwu@bwu.edu.cn,

Wanghan18@163.com

**Abstract.** The concept of intelligent transportation has been accepted by society and has been applied widely. At the same time, the problems of the intelligent transportation are emerging and the Internet of Things (IOT) provides a new direction for its development. The paper first analyzed the concept of IOT, key technologies and system architecture, and then researched the demands of intelligent transportation's function to put forward the structure of overall function of the urban intelligent transportation management based on IOT. Besides, the paper also discussed the key technologies involved in the urban intelligent transportation management and designed a detailed structural model of the intelligent transportation system based on IOT. Finally, the paper forecasted the potentials of China's intelligent transportation system based on IOT.

**Keywords:** intelligent transportation, Internet of Things (IOT), system architecture.

## **1 Introduction**

Internet of Things refers to a network allows a series of intelligent activities like identification, positioning, tracking, monitoring and management by linking devices like RFID, Smart Sense, GPS (Global Positioning System) and 2-D Code, etc. in objects to wireless network via interfaces to endow objects with intelligence, therefore realize the communication and dialogue between human and objects as well as objects and objects.

The application of IOT has been spread from intelligent transportation, logistics and scheduling tracking, and base station's monitoring for business to individual medical treatment and smart home, etc. for public. It has been extended to all walks of life, but it is still in its initial stage and has not been popularized in a large-scale. Intelligent transportation industry always uses IOT technology, networks and devices to achieve

\* Supported by National Natural Science Foundation of China(70031010) , Funding Project for Academic Human Resources Development in Institutions of Higher Learning Under the Jurisdiction of Beijing Municipality(PHR200906210), Beijing-funded project to develop talents(PYZZ090420001451).

intelligent transportation and they have already merged with each other. The industrialization of IOT will vigorously promote the great development of ITS in China. Intelligent transportation industry has been recognized as one of the industries with the most potential to achieve success by utilizing the industrialization of IOT into practice.

ITS applies the sensor technology, RFID technology, wireless communication technology, data processing, network technology, automatic control technology, video detection and identification technology, GPS, and information dissemination technologies to the entire transportation management system so that establish real-time, accurate, and efficient comprehensive management and control system of transportation. The progress of intelligent transportation will drive the development of the smart car, navigation, RFID, the perception technology of operational state of transportation infrastructure (such as smart roads, smart rail and smart waterways, etc.), communication technology between vehicles and transport infrastructure, communication technology between same or different vehicles, and dynamic real-time traffic information dissemination technology. It enjoys a wide range of applications. The paper researches the urban intelligent traffic management system on the basis of IOT.

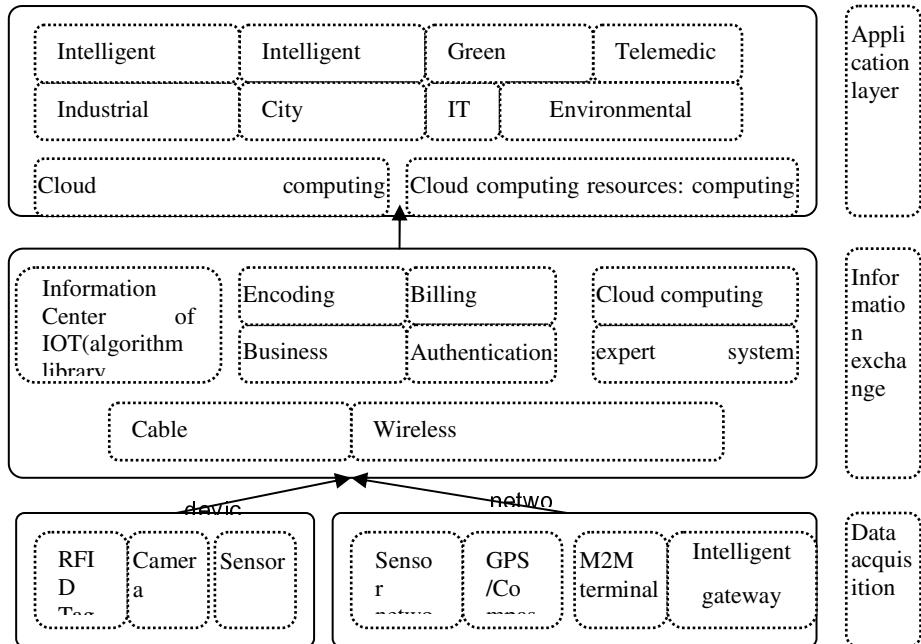
## 2 The System Architecture of IOT Technology

Internet of Things is an open architecture with a variety of technical supports, including radio frequency identification technology, middleware technology, logistics management, e-commerce technology and so on. It involves three key technologies: (1) sensing technology--obtain the object information simultaneously and accurately through the use of RFID, sensor, two-dimensional product code and other equipments and technologies; (2) information transfer technology-- deliver real-time and accurate object information by using the deep integration between a variety of telecommunications networks and the Internet; (3) intelligent processing technology--analyze and sort out large amounts of information and data as well as implement intelligent control of the goods through cloud computing and fuzzy intelligent recognition technology.

With the analysis of key technologies of IOT, the overall technical architecture of IOT should ensure the scale, mobility and safety. The application of IOT is still relatively fragmented in China, and there is no large-scale, systematic development trend, so a scalable and open architecture of IOT should be established in the industrial chain technology as well as application and formation system in ideal aspects in order to break through the barriers of application in large-scale and promote the IOT industry's transition from the start-up period to the growth period. The overall technical architecture of IOT includes data acquisition layer, information exchange layer and application layer. The overall technical architecture of IOT is shown in Figure 1.

Data collection layer consists of two-dimensional code tags and readers, RFID tags and readers, cameras, sensors, GPS, sensor gateways, sensor networks and other

equipment and technologies. In this layer, it mainly solves the issues like data collection and object identification, etc. It is composed by the various types of acquisition and control modules, and the main function is to complete IOT's information perception, data collection and the control of facilities. It is an important foundation of IOT.



**Fig. 1.** The technical architecture of IOT

The layer of information exchange is based on the network of IOT and communication technologies, such as mobile communication network and the Internet, which is a converged network formed by a variety of communication networks and the Internet. It includes information center, management center of IOT, expert systems and cloud computing platform, which are responsible for the massive part of Intelligent Information Processing. Therefore, the network layer not only requires the ability to operate the network, but also the ability to enhance operational efficiency of information. It is the infrastructure to make the IOT become a universal service.

Application layer refers to solutions of integrating IOT technologies with industrial technical expertise to achieve a wide range of application with intelligent technologies. Through the application layer, IOT ultimately realizes deep integration with information technology and industrial professional technologies. The application layer lies on the top which is the ultimate goal of IOT applications. It consists of a variety of servers and its main functions include the collection, transformation and analysis of the gathered data as well as the adaptation and triggers of things for users. The key issue in this layer is socialization of information's sharing and information security.

### 3 The Design Objective and Overall Function of IOT Intelligent Transportation Management System

#### 3.1 Survey of the Demands Urban Intelligent Transportation Management

In the implementation of "The Twelfth Five-Year Plan" of traffic information project in Ji-ning, Shandong, the author launched detailed investigations on the demands of city's intelligent transportation management. The statistical research on demands was done in over 20 subordinate units of Department of Transportation in Jining, Shandong. The author studied the problems from five aspects--government information, data center, transportation web portal, GIS-T and logistics system's construction, with questionnaires and the main elements of traffic information planning and construction. The statistics is shown in Table 1 (A. well-established and run well / B. has been established but has yet to be improved / C. is being established / D. has not been established but needs to be / E. there is no need to establish the system). In the course of the investigation, some respondents did not participate in certain options.

**Table 1.** The investigation on the demands of key transportation management

	A	B	C	D	E		A	B	C	D	E
1.Transit Management Systems	10%	58%	11%	5%	0	16、Data Base					
2.Transportation Management	26%	26%	5%	11%	5%	16.1.Standard Data Base	5%	5%	0%	42%	11%
3.Traffic Inspection System	26%	26%	11%	16%	5%	16.2.Transportation Geographic Information Data Base	0%	16%	5%	37%	11%
4.Road Administration	16%	5%	5%	32%	16%	16.3.Water transport, ports, waterways, and maritime management database	0%	0%	5%	32%	21%
5.Station Monitoring	11%	28%	30%	18%	0%	16.4.Road Transport Management Data Base	16%	0%	0%	5%	0%
6.Online Approval	21%	21%	0%	21%	0%	16.5.Technology Education Data Base	5%	0%	0%	47%	16%
7.Emergency Command	0%	25%	15%	18%	5%	16.7. Policies and Regulations Data Base	0%	5%	11%	47%	11%
8.Charges	26%	42%	0%	5%	5%	16.8. Basic Road Information Data Base	11%	16%	5%	26%	11%

**Table 2.** (Continued)

9.Logistics Stowage	0%	5%	5%	53%	5%	16.9、The Data Base of the city's operating vehicles	11%	16%	5%	26%	11%
10、GPS	11%	21%	0%	26%	5%	16.10. Document Data Base	0%	5%	0%	53%	5%
11.Remote Video Monitoring	10%	15%	15%	12%	5%	16.12. Comprehensive Statistical Data Base	5%	32%	0%	11%	5%
12.Vehicle Management	5%	42%	0%	16%	5%	17. Logistics Information Service Platform					
13.Transportation Web Portal	37%	53%	5%	5%	0%	17.1Logistics Data Exchange Center	68%	0%	0%	0%	32%
14.Date Base of Charges	32%	37%	0%	0%	5%	17.2 Port Logistics Information Service Platform	32%	16%	0%	0%	53%
15.GIS-T	80%	0%	0%	0%	20%	17.3 Road Logistics Information Service Platform	68%	0%	0%	0%	32%

### 3.2 The Overall Function of Intelligent Transportation Management Based on IOT

According to results of investigation on the demands of city's intelligent transportation management coupled with the IOT technology, the functions of intelligent urban traffic management are as follows:

1) *Administration of Intelligent Transportation Management: administration and management of road and waterway transport, port management and maritime management*

The e-government integrated information platform of transportation can provide all kinds of application systems in transportation industry with unified and secure services of authentication and authorization; exchange and share information of transportation with office operations in different industries; as well as bear all kinds of hardware and software platforms for business and information services to provide users with unified services. In the Intranet of government administrations, it can provide internal services like secure authentication, data-sharing, cooperation in business, decision supporting and data mining. While in the outer net, as the uniform external portal of e-government system, it publicizes government affairs to public and offer services like administrative examination and approval, besides, it also provides all kinds of application systems in transportation industry with unified and secure services of authentication and

authorization. E-government integrated information platform will improve the development efficiency of various applications in transportation to meet the requirements of security and confidentiality in office applications of businesses.

E-government is not just to construct a website, the key lies in the collection and sharing of the basic information. Government As a network system to connect different things, IOT meets the requirements of e-government of transportation for basic data to a great extent. IOT handles the collected data coarsely, and then sends the data to the Information Center to process. Taking IOT as the infrastructure of e-government will help to break the barriers between different departments to achieve sharing information among divisions.

*2) Intelligent Transportation: networked charging on highways, IETC, and urban public transport priority, etc.*

At present, intelligent transportation has been applied in many cities which mainly involved in non-stop in freeways and networked charging, multi-path identification in urban roads, as well as urban public transport priority. Many systems related to intelligent transport have been developed and each has its advantages and disadvantages. However, on the whole, the accuracy of forecast, sensitivity of response and adaptability of intelligent transportation system need to be improved. Besides, the collected traffic information is inaccurate, the information feedback is not timely, and the urban traffic environment is complex and variable, all of which restrict the development of intelligent transportation system. Functional structure includes information gathering, information dissemination, and management and control centers, etc. Information collection consists of the ultrasonic microwave vehicle detector installed on the roads, light beacons, image-based vehicle detectors, traffic surveillance cameras and AIV; it also includes gathering traffic and environmental information through satellites, meteorological and environmental monitoring, police and vehicles. Information mainly refers to the feature information of traffic flow (e.g. flow rate, speed and density, etc.), traffic emergency information (various ways to get event information, including: road detector information and labor reporting information, etc.), in-transit vehicles and drivers' real-time information (e.g. location information of vehicle, etc.), environment information (e.g. atmospheric conditions and pollution information, etc.) as well as dynamic traffic control and management information. The information can be disseminated through the traffic signals at intersections (light and sound, etc.) to implement access control, through traffic information board (text and graphics, etc.) to release information, and through the vehicle devices (radio, television and Internet, etc.) to publicize the data.

*3) Logistics and transport: vehicle, container, ship, cargo and yard management, etc.* In logistics and transport, IOT can directly provide services to it. Whatever vehicles, containers or ship management, IOT embeds electronic tags in the relevant objects. For example, bar codes can store objects' identification information and send the instant messaging to the background information processing system by wireless network. Each system interconnects with each other to form a huge network, and thus realize the object tracking, monitoring and other intelligent management in the entire transport projects.

## 4 The System Architecture of Urban Intelligent Transportation System Based on IOT

In terms of traffic information collection, the terminal nodes of intelligent transportation based on IOT utilize non-contact magnetic sensors to regularly collect and perceive the speed of vehicles and distance between vehicles within the region. Many terminal nodes assemble their gathered information and date after the initial processing to the gateway node via sink node to integrate data and access to the road traffic and vehicle speed information, and thus provide accurate input information to the traffic signal control at intersection. Through the installation of sensors like temperature and humidity, illumination, gas detectors at terminal nodes, they can also detect road conditions, visibility and vehicle exhaust pollution. The technologies of intelligent transportation system mainly include electronic information technology, computer technology, digital communication technology and artificial intelligence control techniques, which can be used comprehensively in transportation management, services and control.

Based on IOT, the design of system architecture is divided into sensing layer (data acquisition layer), network communication layer, data integration and computation layer, and application layer. Sensing layer collects traffic information mainly through a variety of sensors, including traffic information collection devices, control systems and information dissemination equipment. Network communication layer transmits and exchanges data by sensor networks, GPS, GSM, ZigBee, and other means of communication and networks, including hardware, network platform and equipment. Data integration and computation layer is responsible for unified standardized format conversion and data integration for collected and accessed data in accordance with uniform data exchange standard. It mainly integrates and processes the basic elements and basic data of traffic management as well as classifies, integrates, formulates tables and database, and constructs data bank of urban intelligent traffic management. The data base is shown in the data center of Table 1.

The application layer provides services directly to the intelligent traffic management, decision-making and the public. Application layer includes traffic information collection system, intelligent traffic monitoring system, public transport management system, the system of rapid response to emergency, electronic police system, traffic guidance system, command and control center, and integrated management platform. Command and control center gathers the traffic tendencies and events of the city to realize real-time linkage and complete intelligent dispatching and other services. The system architecture of urban intelligent transportation system based on IOT is shown in Figure 2.

In addition to technical design and system applications, the intelligent traffic management system should also attach importance to the construction of ITS information security system and operation management security system. Information security system can protect data security and prevent data loss and destruction, which is an important part of intelligent traffic management and one of important measures to strengthen information security. Operation management security system is responsible for providing organizational security and system security for the smoothly operation of intelligent transportation system construction, which includes organizational structure, the construction of qualified personnel as well as operation and maintenance mechanism.

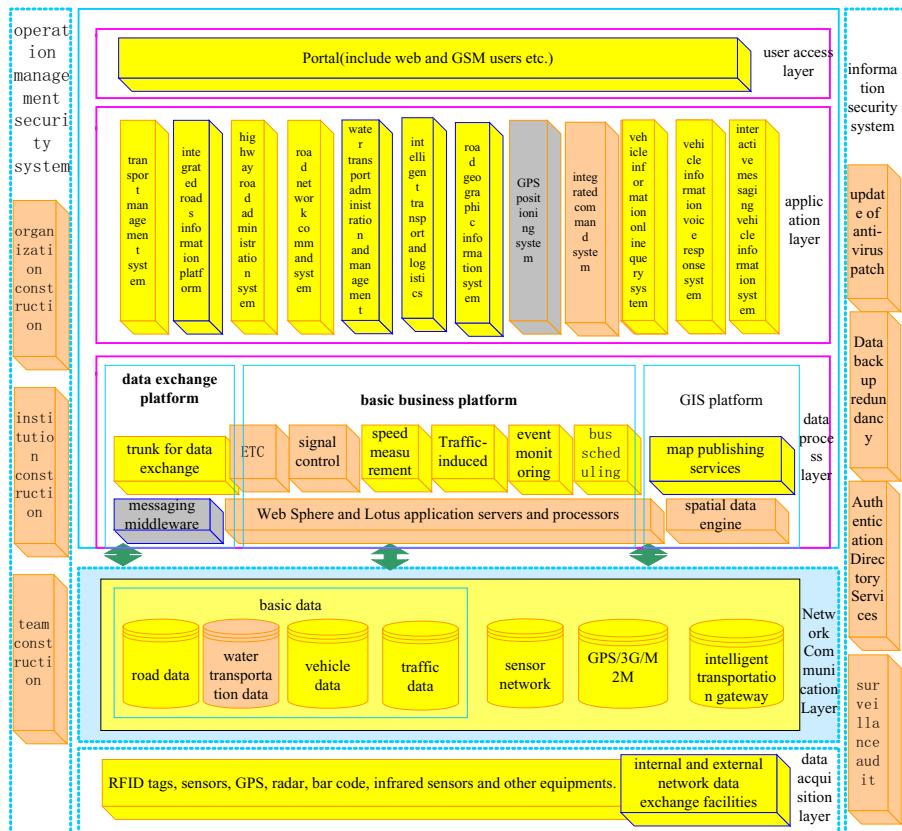


Fig. 2. The system architecture of urban intelligent transportation system based on IOT

## 5 Conclusions

As the development of China's urbanization, the demands of fast and efficient urban intelligent transportation by people are inevitable. The technology levels of intelligent transportation system are similar to IOT's three-layer network architecture. However, the perception of current intelligent transportation system to vehicles is passive, the existing information is incomplete and unshared, and networking and monitoring functions are not strong enough, all of which results in the low integration and low intelligent level. At present, the development of IOT is in its initial stage and the key technology is still at the exploratory stage, but its momentum is strong. With the development of IOT techniques and huge investment in intelligent transportation of the cities, intelligent transportation will enjoy a wide range of industrial applications, and intelligent level of urban traffic will witness a qualitative leap. Internet of Things technology will enable the development of transport system to become more intelligent, safe, harmonious, and energy-efficient.

## References

1. Xie, H., Dong, D.-C., Ou, D.-X.: A New Generation of Intelligent Transportation based on the Internet of Things. *Technology & Economy in Areas of Communications* (01) (2011)
2. Li, Y., Wang, J.-B., Dong, L.-B., Zhou, G.-Z., Song, J.-D.: Research on the application of IOT in intelligent transportation. *Mobile Communications* (15) (2010)
3. Meng, L.: Research and Design of Real-time Intelligent Traffic System Based on ZigBee. *Northeastern University* (2010)
4. Liu, Z.-S., Wei, F., Chai, Y.-T., Shen, X.-S.: Study on the Construction of the Internet of Things in China. *Logistics Technology* (07) (2010)
5. Liu, H.-J.: Research on Key Technology for Internet of Things. *Computer Era* (07) (2010)
6. He, K.: The Key Technologies of IOT with Development & Applications. *Radio Frequency Ubiquitous Journal* (1) (2010)
7. Jiang, Z.-C., Cheng, G.-T., Zhao, Y.: Discussion on Intelligent Traffic System Construction. *Journal of North China Institute of Aerospace Engineering* (2010)
8. Cui, Y.-S.: Research on application of city intelligent traffic integrated platform. *Information Technology* (2010)

# A Single Sign-On Scheme for Cross Domain Web Applications Based on SOA<sup>\*</sup>

Enze He and Qiaoyan Wen

State Key Laboratory of Networking and Switching Technology  
Beijing University of Posts and Telecommunications,  
100876 Beijing, China  
[{heenze,wqy}@bupt.edu.cn](mailto:{heenze,wqy}@bupt.edu.cn)

**Abstract.** The SSO (Single Sign On) is one of the most popular enterprise business integrated solutions. The SSO means that users could only login once to access all the mutual trusted applications. The existing SSO schemes lead into much modification to original system when adding SSO to new application, which means high coupling relation between applications. In this paper, we proposed a SSO scheme based on SOA which would make business system, authentication proxy, and authentication authority management as separate services. The proposed scheme uses enterprise service bus (ESB) to accomplish information interaction, ticket transmission and implement cross domain SSO. The login authentication of business system cloud use authentication proxy to realize different forms of authentication. The scheme is based on PKI/PMI, which achieves strong identity authority and flexible permission management. The result shows our scheme is a high secure, broad perspective solution to the problem of high coupling in SSO.

**Keywords:** single sign-on, SOA, cross-domain authentication.

## 1 Introduction

With the continuous development of information technology, various enterprises and institutions have established a variety of Web applications according to their needs, such as office automation systems, customer relationship systems, financial management systems, etc. These applications have their own separate authentication mechanisms. When a user switches between different systems, entering the corresponding account number and password each time to the user's actions caused a lot of inconvenience. Therefore it needs a unified login system, which users need to link the various application systems. In view of this, the concept of single sign-on came into being. The concept of single sign-on is that when users get authentication in a login point can access a group of the applications associated with the login point

\* This work is supported by NSFC (Grant Nos. 60873191, 60903152, 61003286, 60821001), the Fundamental Research Funds for the Central Universities, (Grant Nos. BUPT2011 YB01, BUPT2011RC0505).

[1]. When logon on one business system, users can access the other business systems that give users the authority. The substance of single sign-on is security context or credential's transferred and shared between multiple applications [2]. It can implements that users can access the resources of multiple applications that gives users the authority by take the initiative to conduct a process of authentication, without their active participation in the subsequent authentication process, to achieve the purpose of "single sign-on, everywhere access".

Currently, there are already many protocols to support single sign-on, which typically have YALE-CAS, Microsoft's PASSPORT, SAML and Liberty Alliance's Liberty [3].

With the emergence of new technologies like SOA, PKI, IBC [4], the single sign-on technology began to develop to the direction of standardization.

SOA is the abbreviation of Service-Oriented Architecture, the basic idea is to serve as the core, the enterprise IT resources integrated into operational, standards-based service, it can be re-combined and applied. With SOA be used for single sign-on program design, will greatly reduce the system's coupling, and improve the system's reuse.

PKI and PMI are dealing with the two functions of user identity and user authorization separately. Therefore, we can use PKI and PMI provides encryption, signature, integrity, non-repudiation, flexible authorization and other security services for the design of single sign-on system, so that the whole system will be more comprehensive and secure.

## 2 Related Work

Reference [2] proposed a Scheme which introduces a concept named certification audit authentication middleware which does work of connection among Login page, authentication proxy, authentication and authorization system. This scheme's shortage is that users have to rely on a plug-in named cross-domain IDP to achieve cross-domain. Multiple domain systems, the system's coupling is high.

Reference [5] proposed a PKI-based web single sign-on solution, using two-way authentication, two-stage authorization and SSO proxy, but did not do research on cross-domain issues.

Reference [6] considered Windows Passport had obvious defects in usability, compatibility, security, and the scheme required the user to centrally stored in the Microsoft Certification was a major flaw that could lead to mutual authentication with other authentication systems failed; Liberty was mainly limited to user account password authentication.

Reference [7] designed a SOA-based cross-domain single sign-on system, reflected the nature of the characteristics of loose coupling in SOA, but there are two problems: (1) A single authentication method, only the user account and password authentication, cannot meet the diverse certification requirements. (2) Single sign-on system requires high security, this system only relies on SSL protocol which is the secure channel between the HTTP and IP layer, the token time-out failure, and immediately destroy the token after the verification to ensure safety. This scheme did not specifically discuss the security issues.

In summary, the existing single sign-on scheme neither solved the high coupling problems between cross-domain systems, nor the decoupling method is common, such as relied on the media dependence. Some schemes can effectively solve the problems of cross-domain and high coupling, but the system's security is not strong. Therefore, a good single sign-on scheme should consider coupling, reusability and security at the same time.

This paper presents a new single sign-on scheme, which is based on SOA. It also uses PKI/PMI as authentication and authorization mechanisms, not only has the advantages of loosely coupled, reusable, platform-independent, etc. but also has a high security, by using the multiple authentication modes to meet the security requirements of different applications.

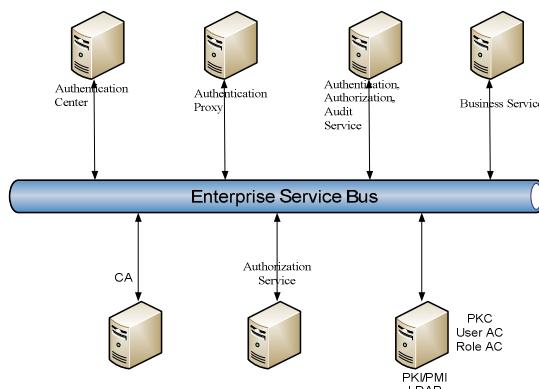
### 3 Scheme Description

#### 3.1 Scheme Architecture

The program is based on SOA architecture with combination of PKI / PMI. By taking the advantages of SOA, such as loosely coupled, platform independence, reusability, we make authentication proxy, authentication and authorization management module, and business system as service to provide a unified interface to the outside world in order to add more flexibility to the system. Using services that PKI / PMI provides such as encryption, signature, integrity, non-repudiation and other security services to increase the security of the system.

In this scheme, security credential is stored in the client, while more than one business system to share information of the credential, to ensure that users properly log on or off multiple business systems. This service applies only to B / S model, so that the business uses cookie to store the ticket and other information.

The overall architecture of single sign-on scheme based on SOA is shown as figure 1:



**Fig. 1.** The overall architecture of single sign-on scheme based on SOA

Functions of the architecture are as follows:

- (1) **Authentication Center:** Authentication center provides the login portal and login authentication for users of the mutually trusted systems. Login portal provides a unified web login page for the user, and call authentication authorization audit service for identity authentication. After authenticating users successfully, the system uses Cookie or Session to store user information. User chooses to access an optional service after login successfully. Authentication Center uses a URL to redirect to the selected system which returns a ticket to the user. Authentication Center verifies whether the user has been logged by checking Session or Cookies.
- (2) **Authentication Proxy:** The function of authentication proxy is providing multiple authentications for users, for instance: user name / password, USBKey, fingerprints, digital certificates and so on. Users can choose the login authentication on the login web page.
- (3) **CA:** User will apply certificate during registration. Each user will be bound with a PKC which has a unique certificate serial number. CA provides encryption, signature, integrity, non-repudiation and other services for the whole system.
- (4) **Authorization Service:** Authorization service which is based on PMI provides a unified interface to the outside world and implements the authorization based on RBAC (Role-Based Access Control). PMI issues User AC (user attribute certificate), Role AC (role attribute certificates) that are stored in LDAP (Lightweight Directory Access Protocol) for user.
- (5) **PKI/PMI LDAP:** Store the PKC and AC that are issued by PKI/PMI.
- (6) **Business Service:** Business service that users want to access uses Web Service which provide unified interface to the outside world connects to the Enterprise Service Bus for information interaction.
- (7) **Authentication Authorization Audit Service:** This service provides unified interface to outside and connects to Enterprise Service Bus .The service mainly achieves authentication, authorization and audit functions, and according to certain strategies to achieve the user access control. The main function is as follows:
  - a) Authentication: The service can support a variety of login authentication function. On B / S services, user access to the unified Web Authentication page and deal with client events through Activex control, such as the electronic key certificate acquisition and so on.
  - b) Authorization: System depending on the user to grant access to different services, this service is responsible for determining the user's permission.
  - c) Audit: System stores the user's operating records.

### 3.2 The Process of Cross-Domain Single Sign-On

Single sign-on domain is the domain of service entities group which has one trusted authority which is trusted by all the service entities in this domain [6]. As the

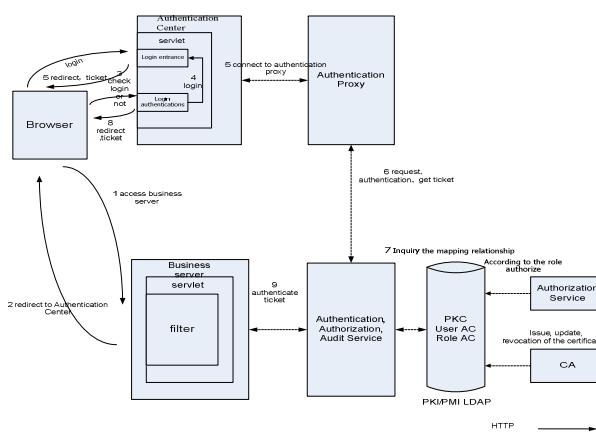
restriction of cookie's properties makes it available only for the hosts in the same domain, but distributed applications often cannot guarantee that all the hosts in the same domain. In this scheme, both local services and cross-domain services put the ticket information as the parameters in the URL, the business service invoke the appropriate Web Service interface to complete the authentication of ticket. The AAA (Authentication, Authorization, Audit) Service, which is composed of AS (Authentication Server) and TS (Ticket Server), play a key role in the process of authentication and ticket building. AS and TS are trusted between each other. CA issues X.509 digital certificates to AS and TS.

The identifiers used in this paper are as follows in table 1:

**Table 1.** Identifier

Identifier	Implication
$PKC_x$	The digital certificate of X
$KU_x$	The public key of X
$KR_x$	The private key of X
$E_x[P]$	Use the key of X to encrypt the plaintext P
$D_x[P]$	Use the key of X to decrypt the ciphertext P
$Hash[M]$	Do the Hash operation to message M
$K_c$	The key that AS used to encrypt cookie
$IP_n$	The IP address of users

The single sign-on process when user accesses the business service without being authenticated by AS is as follows in fig. 2:



**Fig. 2.** The process of cross-domain single sign-on

- 1) Users require visiting Business Server.
- 2) Business Server discovers the user has not been authenticated by AS, Thus, redirects user to Authentication Center. If user has been authenticated by AS, the process will go to 8).
- 3) The login authentication model of Authentication Center checks user's session or cookie information, if user has not been logon, the system will redirect user to login entrance and the process will go to 4), else the process will go to 5).
- 4) User logs in and fills the login information according to selected authentication model.
- 5) Connect to Authentication Proxy which will choose different authentication model to authenticate user's identity in the Authentication Authorization Audit Service according to authentication model user has selected.
- 6) AS in The Authentication Authorization Audit Service checks the HTTP request, if the value of cookie in the local domain is null, it means the user has not been authenticated. Users use their own digital certificate PKCU for authentication. AS uses a user certificate serial number pkcID to check PKI / PMI LDAP to get user's global identity in the system named sso(userID), user information and other identity information named userInfo, generates the user's authentication token named userToken. userToken=sso(userID||pkcID||us-Info||IPn||createTime). Produce the token's unique identifier named tokenID which is 24-byte random number through the Secure Cookie [7, 8, 9] sent to the user's browser cache.

Cookie's value is Value=  $E_c [[\text{tokenID}||\text{IPn}]||E_{KR_{AS}} [\text{hash}[\text{tokenID}||\text{IPn}]]]$ .

- 7) AS uses pkcID to query PKI / PMI LDAP to get the corresponding RoleID which is the unique identifier of Role AC. According to the role to determine whether the user has permission to access the web application and then get the user's unique identifier named permission\_roleID to the web application.
- 8) AS requests TS generates ST(Service Ticket), ST=sso(userID||serviceID||IPn||permission\_roleID), TS returns ST\_ID to AS. ST\_ID is a random string of 24 bytes which is used to find ST.
- 9) AS uses URL redirection to direct users to the business server through put ST\_ID as parameter in the URL.
- 10) Business server receives the ST\_ID, request TS authentication. The content of request message is: ST\_ID||serviceID||IPn.
- 11) TS uses ST\_ID to index ST. TS compare the content in ST with the content of the Business Server's request message to check whether they are the same. After authenticating successfully, TS extracts permission\_ID from ST. Permission\_ID is the unique identifier of user in the web application. The content of the response to Business Server is  $E_{KU_s} [\text{permission\_roleID}]||E_{KR_{TS}} [\text{hash}[\text{permission\_roleID}]]$ , KUs is the public key of web applications. For the authenticated ST, TS delete it from the system immediately.
- 12) Business Server using the private key to decrypt the message to get permission\_roleID, and using the TS's public key to verify the digital signature of permission\_roleID. Then it establishes a session with the user and stored ST\_ID in the session.

## 4 Scheme Analysis

### 4.1 System Analysis

This scheme implements SSO based on the combination of SOA and PKI/PMI, to meet the purpose of loosely coupled, platform-independent, reusable and improve the security of the system.

This scheme uses SOAP as the standard protocol for cross-domain authentication and authorization message transfer. Using WSDL to describe, and through the UDDI to register [10]. The biggest feature of SOA is cross-platform, applications in each domain invoke the interface of Web service for data exchanging. According to the feature and the configuration of the domain trusted can implements SSO.

In addition to PKI / PMI, different authentication system, authorization system, and Authentication Authorization Audit service access to enterprise service bus can implement SSO system.

Authentication Authorization Audit Service can support different authentication methods by extending authentication proxy. Different authentication model work as plug-in can support this authentication method by configuration.

The comparison of system architecture with the other references is shown in table 2:

**Table 2.** System architecture comparison

Scheme	System Architecture
Reference[2]	Cross-domain relies on IDP, the coupling is high.
Reference[5]	Insert SSO proxy into web application, large changes, high coupling.
Reference[6]	Authentication proxy works as plug-in, high flexibility.
Reference[7]	Based on SOA ,loose coupling, high reusability.
Reference[10]	Base on Web Service ,loose coupling.
This scheme	Based on SOA ,loose coupling, high reusability.

### 4.2 Security Analysis

- 1) Authentication Security: This scheme uses the authentication based on digital certificate, the legitimacy of users and servers can be guaranteed.
- 2) Cookie Security: After successful authentication, the server sends to the user's browser cache encrypted and signed Secure Cookie. User's cookie data is encrypted with different keys to ensure message's confidentiality, and signed with the server's

private key. Any revision of the cookie data, Authentication Authorization Audit service can check it out.

Using server key to do HMAC operation to user's certificate serial number and expiration time will do good. The values to be encrypted are different each time.

3) Defense MITM (Man-in-the-Middle-Attack): This scheme encrypts and signs key information (such as cookie information, shared memory information), while the entire messaging process uses SSL encryption and defense MITM effectively.

4) Ticket Security: ST (Service Ticket) which contains user authentication and authorization information is stored in TS cache. Index number ST\_ID which is transferred in network is a long random string. Attacker intercepts ST\_ID cannot get any useful information. And ST is designed to be disposable, is removed immediately after successful authentication. If the ST\_ID which is requested by business server for authentication is invalid, it can be considered to be the replay of ST\_ID. The system will not do any response to reduce system's expenditure.

The comparison of system security with the other references is shown in table 3:

**Table 3.** System architecture comparison

Scheme	System Architecture
Reference[2]	PKI/PMI, Secure Cookie, Multiple Authentication Modes. Strong security.
Reference[5]	Based on PKI, mutual authentication, Secure Cookie. Strong security
Reference[6]	The protection mechanism for cookie and authentication is not perfect, security is not strong
Reference[7]	No research in security, the system security is weak.
Reference[10]	Use user name and password only for authentication, the system security is weak.
This scheme	Based on PKI/PMI, Secure Cookie, Multiple Authentication Modes.Strong security

## 5 Conclusions

In this paper, we proposed a new single sign-on scheme which is based SOA combined with PKI/PMI. SOA has the advantages of loosely coupled, reusable, platform-independent, etc. Meanwhile, based on PKI / PMI using signatures, encryption and other techniques, can effectively prevent content tampering, replay attacks, middle attacks, phishing and other attacks. The scheme is compatible with a variety of business systems and different authentication methods, and supports cross-domain, can be used in e-government, e-commerce, digital campus and other areas.

## References

1. Manshan, L., Heqing, G.: The present situation and development of single sign-on technology. Compter Applications 24, 248–250 (2004)
2. Li, X., Wen, Q., Dai, Z.: A Supporting Multi-Mode Application Single Sign-On Scheme Based on PKI/PMI. Journal of Beijing University of Posts and Telecommunication 32(3), 104–108 (2009)
3. Wang, Z.: Study and implementation of cross-domain single sign-on system based on Java EE platform. Southwest Jiaotong University, CA (2010)
4. Wang, Y., Wen, Q., Zhang, H.: A Single Sign-On Scheme For Cross Domain Web Applications Using Identity-Based Cryptography. In: 2010 Second International Conference on Networks Security, Wireless Communications and Trusted Computing (NSWCTC 2010), pp. 483–485. IEEE Press, Beijing (2010)
5. Shi, R., Xu, L.: A PKI-based Single Sign-on Scheme for Web Application. Micro-Computer Information 26(3), 39–41 (2010)
6. Huang, C., Li, Z., Yang, Y.: A New Web Single Sign-on Scheme Supporting the Multiple Authentication Modes. Journal of Beijing University of Posts and Telecommunication 29(5), 130–134 (2006)
7. Niu, X.: The Research and Implementation of a SOA-based Cross-Domain Single Sign-On System. Beijing Jiaotong University, CA (2009)
8. Liu, A.X., Huang, C.T., Gouda, M.G.: A secure cookie pro-tocol. In: Proceedings of IEEE ICCCN 2005, pp. 333–338. IEEE Press, San Diego (2005)
9. Park, J.S., Sandhu, R.: Secure cookies on the Web. IEEE Internet Computing 4, 36–44 (2000)
10. Liang, Z.: The present situation and development of single sign-on technology. Compter Applications 30, 3363–3367 (2010)

# A Fast FCD-Based Dynamic Traffic Navigation System

Shusu Shi, Zonghui Wang, Licheng Yu, and Wenzhi Chen

College of Computer Science and Technology, Zhejiang University

310027 Hangzhou, China

{shusushi, zjuzhwang, yulicheng, chenwz}@zju.edu.cn

**Abstract.** FCD-based dynamic traffic navigation system, which takes effective traffic management methods, is more and more popular in the world. This paper analyses techniques of a FCD-based dynamic traffic navigation system and presents the framework of the system. Then, the paper introduces its implementation. After that the prototype system is tested and found that the system can handle the large-scale floating car data quickly and effectively, and responds well to real-time traffic information.

**Keywords:** Floating Car, Constrained A\*, Dynamic Traffic Navigation.

## 1 Introduction

In the progress of urbanization, the existing traffic capacity can't meet the growing demand in many cities. Traffic jams are more or less becoming big problems in the world[1]. Therefore people began to study more effective traffic management methods. ITS(Intelligent Transportation System) is one of them.

The most commonly used application in an ITS is dynamic traffic navigation. Real-time traffic information can be collected either by installing monitoring equipment or by collecting FCD(floating car data). While installing monitoring equipment may lead to high cost but low coverage rate when the map is huge, collecting FCD and dealing with it is a more sensitive approach.

This paper researches the FCD-based dynamic traffic navigation system and implements a prototype system which we found can deal with FCD and push real-time traffic information quickly after testing it.

## 2 Related Works

The first problem of FCD-based traffic computing is how to match the location of a floating car to road network. The basic approaches, which are classified by Sun Dihua [2], are to match the locations from point to point, from point to curve, and from curve to curve. Christopher et al, and Li Yuguang proposed a weighted metric method[3,4]. Christopher et al also proposed the use of road constraints to help road-matching[3]. Syed et al and Rajashri proposed the methods that need some other

expertise [5, 6, 7]. Among all the approaches mentioned above, the point-matching method is the simplest. But it's effectiveness is relatively poor and is dependent on the storage of road network. Point to line mapping method is more commonly used[1] and much more effective, but it is not stable when there are intersections or when the road network is complex[4]. Curve to curve approach needs to define a suitable distance between the curves[3], and this method requires relatively a large amount of computation. Another problem of FCD-based traffic computing is that before calculating the speed of the vehicles between two floating cars' location, we need to find out the shortest path between them. In the dynamic traffic navigation, the shortest and fastest path searching are also the core parts. The classic approaches include Dijkstra algorithm[8] and Peter's A\* algorithm[9]. However, these two classic algorithms do not support to search on road network with constraints. Existing algorithms which support the constraints such as Li Yinzhen's and Ku Xiangyang's both require the road network itself to be modified before calculation. Besides, they're both based on the traditional non-information searching method, so the calculation speed is not high.

### 3 Analysis

The implementation of FCD-based dynamic traffic navigation system is mainly based on solving three problems: road matching, shortest(fastest) path searching, and traffic speed estimation.

#### 3.1 Road Matching

This paper takes point to curve method as the basic approach of road matching. This method is simple, easy to be implemented, and efficient. Therefore it is often used in practice[1]. But when there are intersections or when the point is near separated two-way roads, it easily maps to a wrong road. In the implementation, in order to improve efficiency, we use an approach called map blocking to speed up the matching. Map blocking divides the map into several uniform blocks, checks which block the floating car is located in, then the matching will be constrained among the roads in the block. It also takes blocks around into consideration in case that the floating car is located near the edge of a block, which may make the matching wrong.

#### 3.2 Shortest Path Searching

A\* algorithm with constraints is used for the shortest path searching in this paper. The main difference between A\* algorithm with constraints and the classic one is that, the former takes constraints R, such as one way street, intersection turning prohibitions into account, where  $R = \{r_i (v_{i1}, v_{i2}, v_{i3})\}$ ,  $v_{ij} \in V$ . An element  $r_i$  in R is a rule which represents that it is prohibited to go from  $v_{i1}$  to  $v_{i3}$  through  $v_{i2}$ .

The method of A\* algorithm with constraints to meet actual demand is to check whether it can arrive  $v_i$  through one of node  $v$ 's parent nodes,  $v_p$ , when the original A\* algorithm expands the node  $v$  out to a child node  $v_i$ . If for all nodes  $v_p$ , there is  $r(v_p, v, v_i) \in R$  exists, then it's believed that node  $v_i$  is unreachable from node  $v$ .

### 3.3 Traffic Speed Estimation

After the match, the traffic speed of a road could be estimated. For each floating car, the system finds out the travel path from the former mapping point to current mapping point, and calculates the speed in accordance with the time difference. For each road section, the system defines the estimated speed as the average speed of all vehicles on the road.

## 4 Framework and Prototype System

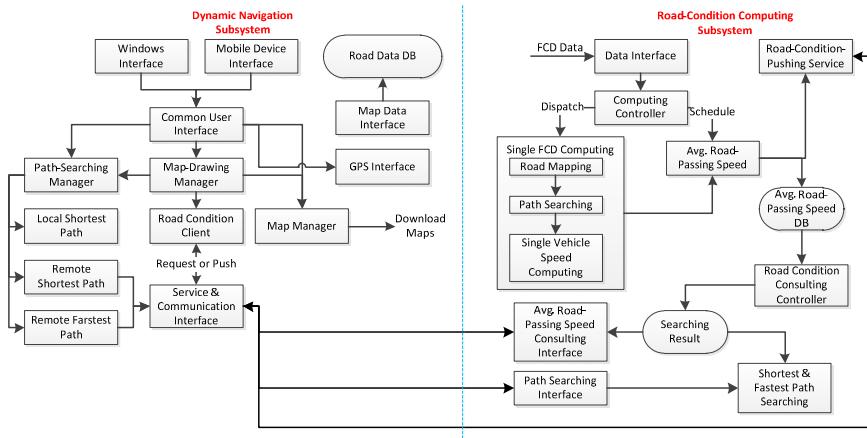
### 4.1 Framework

The framework of the FCD-based dynamic traffic navigation system is shown in Fig. 1.

In Fig.1, Data Interface receives real-time FCD, and filters the data according to its effectiveness.

In Road Condition Computing Subsystem, computing controller is the core of control, and its main function is to distribute data and to schedule tasks. Computing of single FCD is completed by Single FCD Computing module, and the result is used as input of Avg. Road-Passing Speed module. After the estimation of average traffic speed, the result is written into Avg. Road Passing-Speed Database, and at the same time it is published through Road-Condition-Pushing Service. Other operations are initiated by the client that links to the subsystem's service interface.

Dynamic Navigation Subsystem is implemented on mobile devices. In the subsystem, the Map Drawing Manager module is responsible for the management of drawing of all maps in the subsystem. Path Searching Manager provides a common representation for paths, and it can call different path searching modules so that different algorithms could be used. Road Condition Client module saves the traffic information that obtained from road condition server. In the meantime it caches and manages the data. Service & Communication Interface is used to communicate with road condition query service or publishing service. It is used to be called by remote path searching module or client module. Map Data Interface provides a uniform representation of map data for the convenience of changing map. GPS interface receives GPS data from a device's serial port, then analyses it, and shows the location on the map.



**Fig. 1.** Framework of the FCD-based dynamic traffic navigation system

## 4.2 Prototype System

We finally implemented the prototype system according to the framework in Fig. 1.

Road condition computing subsystem runs on high-performance dedicated server. It is implemented in Java.

Dynamic navigation subsystem is implemented on Windows, Windows Mobile, Android and iOS platform.

Fig. 2 (a) and (b) show the result of shortest path searching and road traffic information on the map respectively.



**Fig. 2.(a)** The result of shortest path searching



**Fig. 2.(b)** Road traffic information

## 5 Result

This section tests speed estimation accuracy and data processing performance of the prototype system.

All tests in this section are on the PC, with 3.06GHz Intel Core 2 Duo T9900 CPU, 8GB memory, Microsoft Windows 7 Professional x86\_64, and Java 1.6.0 Update 18 runtime environment.

### 5.1 Speed Estimation

The speed estimation accuracy is measured by comparing the estimated speed with real-time traffic speed which is obtained by driving along the road and the comparison system (<http://www.hzctrc.com/hzgis/bin-debug/TrafficMap.html>), which is developed by Hangzhou City Transportation Research Center. The comparison system is based on the same input as the prototype system.

**Table 1.** Speed estimation of the two systems' and real-time speed

Road	Real-Time	Speed(km/h)		Road	Real-Time	Speed(km/h)	
		Prototype System	Comparison System			Prototype System	Comparison System
1	30	16.1	-	9	20	25.6	-
2	30	28.9	-	10	<10	20.4	13.4
3	30	27.8	29.15	11	40-50	45.2	-
4	20-30	20.8	23	12	40-50	46.9	41.1
5	20-30	21.9	26.1	13	40	41.9	-
6	10	19.9	16	14	30-40	12	-
7	30-40	15.8	16	15	30	34.2	33.65
8	12	15	19.42				

As shown in table 1, on the road 2, 3, 4, 5, 11, 12 and 13, the prototype system gives relatively accurate results. These roads are characterized by simple lines, less traffic lights, even traffic speed, and a large number of floating cars.

However, the prototype system gives inaccurate speed results on road 6 and 7. These two are characterized by a long-time red light.

Compared to the comparison system, the prototype system gives a more smooth estimated speed. Due to the fact that comparison system sets the time interval of updates to 5 minutes, while the prototype system sets it to just one minute, the traffic lights have a greater impact on the prototype system.

### 5.2 Performance

This section evaluates the data processing performance of the road condition computing subsystem. There are about 5,000 floating cars and they return FCD every 30 seconds. Therefore, the duration of 3,000,000 records are about 5 hours. The time consumed by the subsystem to handle the data is shown in Table 2.

**Table 2.** Time Consuming

Total Time	Speed of Handling (records/s)		
	All Records	Efficient Records	Mapped Records
18 minutes, 52 seconds	2650.2	2414.9	1902.3

The total time includes estimating the average speed. Since all the records are handled continuously, it shows that 5-hour-record can be handled in just 18 minutes and 52 seconds. It can be inferred that the system can handle at least more than 1900 records per second. The number of floating cars in Hangzhou now is about 8,000, so the actual need of the average processing capacity is 267 records per second considering the fact that the cars return records every 30 seconds. Obviously the current processing capacity for the system far exceeds the actual demand.

## 6 Conclusions

FCD-based dynamic traffic navigation system is an important part of ITS. This paper presents the framework of a FCD-based dynamic traffic navigation system and introduces its implementation. After that the prototype system is tested and we find that the system can respond well to real-time traffic information, and can handle the floating car data quickly and effectively.

The accuracy of speed estimation still can be improved for better performance. Moreover, the computing model should be improved to get more accurate results, and to make the system more practical.

## References

1. Dong, J.: Study on link speed estimation in urban arteries based on GPS equipped floating vehicle. Master's degree paper, Chongqing University (2006)
2. Sun, D., Zhang, X., Zhang, Z.: Map matching technology and its application in ITS. Computer Engineering and Applications 20, 225 (2005)
3. White, C.E., Bernstein, D., Kornhauser, A.L.: Some map matching algorithms for personal navigation assistants. Transportation Research Part C: Emerging Technologies 8, 91–108 (2000)
4. Li, Y., Xiong, P., Yue, Y.: Vehicle Travel Speed Estimation from Large Volume Floating Car Data. A Case Study of Wuhan, Journal of Transport Information and Safety 4, 26 (2009)
5. Syed, S., Cannon, M.E.: Fuzzy Logic Based-Map Matching Algorithm for Vehicle Navigation System in Urban Canyons. In: Proc. Natl. Tech. Meet. Inst. Navig. (ION 2004), pp. 26–28 (2004)
6. Joshi, R.R.: A new approach to map matching for in-vehicle navigation systems: The rotational variation metric. In: IEEE Conf. Intell. Transport Syst. Proc., ITSC, pp. 33–38. Institute of Electrical and Electronics Engineers Inc. (2001)

7. Joshi, R.R.: Novel metrics for map-matching in in-vehicle navigation systems. In: Proc. Intell. Vehicle Symp., p. 36. IEEE (2002)
8. Dijkstra, E.W.: A note on two problems in connexion with graphs. In: Numerische Mathematik, pp. 269–271. Springer (1959)
9. Hart, P.E., Nilsson, N.J., Raphael, B.: A Formal Basis for the Heuristic Determination of Minimum Cost Paths. IEEE Transactions on Systems Science and Cybernetics, 100–107 (1968)

# **Design of Field Information Monitoring Platform Based on the Internet of Things**

Keqiang Wang and Ken Cai

School of Information  
Zhongkai University of Agriculture and Engineering  
Guangzhou, China  
wangkq@zhku.edu.cn

**Abstract.** This paper introduces developed Internet of things Farmland monitor platform by wireless sensor network technology and computer network technology. Remote monitoring of crop forecasting and early warning can be implemented by monitor platform. Real-time data acquisitions and real-time communication platform features provide great convenience to peasants and agricultural researchers. Agricultural land based information monitoring platform will be introduced in this paper, such as Internet of Things applied to agriculture, development of agricultural modernization.

**Keywords:** Internet of things, information monitoring, remote monitoring, agricultural information.

## **1 Introduction**

The Internet of Things has first been used in 1999. People can easily guess the meaning of its name, which is uniquely identifiable objects representations in an Internet structure [1,2]. The Internet of Things makes full use of next-generation IT technologies among the industries. Specifically, sensors and equipments are installed in the power grid, railways, bridges, tunnels, roads, buildings, water systems, dams and gas pipelines. All of the objects will be connected to existing internet. Human being integrates Physical system by the Internet of Things. A super -computer cluster is generated when the Internet of Thing is integrating. Human beings, machines, equipments and infrastructures are controlled by super-computer cluster in real time. Based on that, people can be more refined and dynamic management of production and life intelligently, improve resource utilization and productivity levels, and improve the relationship between human beings and nature.

This paper introduces developed Internet of things Farmland monitor platform by wireless sensor network technology and computer network technology. The acquisition of various data is recorded to the database and is showed by table and diagram. The records of crop contribute in early warning and forecasting. These capabilities will bring new agricultural technology changes, and solve the limitations of modern agricultural research, such as the limitations of monitoring sites, the limitation of monitoring time, and the limitations of technology communication.

## 2 Related Technologies

WAMP is a development platform based on Windows, apache, mysql and php techniques. Hypertext Preprocessor (PHP), Php is a script language that runs on server. Mysql is a open source database. Apache is world class web server. It is a simple, high efficiency, and stable server provider. Phpmyadmin is a written by php. It can fully control servers through web and mysql database [3].

Model View Controller (MVC) is software architecture. The pattern isolates “domain logic” from user interface, permitting independent development, testing and maintenance of each.

ThinkPHP is a fast and simple OOP PHP MVC framework [4].

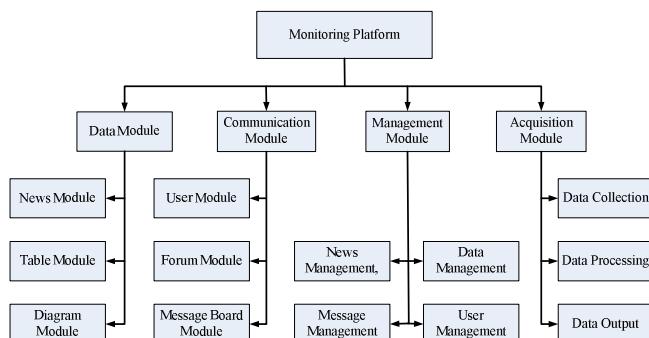
WamServer is a windows web development environment. It packaged Apach, PHP and MySQL.

The existing backbone network communications and special safety requirements of independent access gateway developed by interconnect structure. 32-bit ARM processor, embedded real-time Linux system, 10Mb/s Ethernet, and Wi-Fi interface are used in this design. Wireless sensor networks interconnect with industrial Ethernet can be solved in this design. As a result, this design has a good promotion prospects and great market.

## 3 Overall Design

### 3.1 Framework

Platform frame structure is shown in Figure 1.



**Fig. 1.** Platform Structure

### 3.2 Platform Module Analysis

#### 3.2.1 Data Module Analysis

News module, table module, and diagram module are included in the platform. Updated in real time data capabilities is supported in table module and diagram.

News module includes homepage general news, news, and news details. Furthermore, news submission function is included in news module also. News can be displayed on the web after the manuscript reviewed.

Crop temperature and liquid level is displayed in Chart module. User can directly monitor the environment temperature changes of crop, crop growth and the environment.

In order to delivery PH value of crops and humidity changes directly, graph module will be used in this module.

### **3.2.2 Communication Module Analysis**

Communication Module includes user module, forum module, and message board module. User module provides registration functionality. Users can submit news after login. A broad communication platform is provided by this forum module. Members can post, replies, resource sharing, and modifying personal information in this forum. Convenient communication provided by message board. Users can forward their opinions, suggestions and questions.

### **3.2.3 Management Module Analysis**

Management Module includes news management, message management, data management, and user management. Management is a critical module since site co-ordinating all of the data, only the administrator privileges to operate on the data. News management module can add, edit, delete, and press releases functionality. Message management can capture incoming data and display data. User management can remove users and modify the permissions. Data module can modify the way of display data, such as histogram, graph, and pie chart. Moreover, data module supports data modification, deletion, and backup.

### **3.2.4 Acquisition Module Analysis**

It includes data collection, data processing, and data output. Hardware need to support data collection that is sent by an array. Data collection can be set time interval of data collection in order to make a measurement between server load and real-time collection. There is a filter in the data collection. Valuable data can be selected and moved to database. Data output mainly access and read data from database. Data will be converted into diagram and table and delivered to users.

## **4 Design and Database**

MySQL has a series of excellent performance, such as reliable, ease of use, and scalable. Furthermore, MySQL supports multiple users' access simultaneously, adaptive memory, extendable and maintainable.

Database Normalization requires stability, security, and good capacity in high load data. Database Normalization requires designers deep understand clients' requirement and design suitable logical structure in order to improve efficiency and reduce data redundancy.

Based on requirement analysis, a database called think\_db was built. There are many tables in the think\_db.

- News table: titles, authors, resources, time, and news' body.
- think\_form: ID, email, title, time, news' body.
- Think\_data: time, temperature, humidity, PH value, surface.
- Think\_member: user name, user rights, password, email, and create time.

## 5 Design and Implementation

### 5.1 The Front Modular Design

Homepage news data core code is shown in Fig.2.

```
$list = $News->order('news_id desc')->limit('12')->select();
```

**Fig. 2.** Homepage news data core code

News module read think\_news, and display twelve data on homepage.  
Message board data core code is shown in Fig.3.

```
$data = $Form->order('id desc')->limit(5)->findAll();
```

**Fig. 3.** Message board data core code

Message board data controller read think\_form table, and display five data on homepage.

### 5.2 Forum Module

Field information, which is a professional website for agricultural technology theme, is developed by Discuz7.2. It provides convenient to users and real-time exchange of technology services. Forum divided into three main sections: Resources zone, technical zone, information zone. The most important one is technical zone. Technical zone divided into two sections also: technology-sharing and technical Q&A. Users can share with their techniques and experience on technical zone.

### 5.3 Data Module

Data module is the critical module, which includes temperature, humidity, PH value and surface. Due to limited conditions, simulation methods will be used for data collection and data output.

Fig.4 shows the example of data collection.

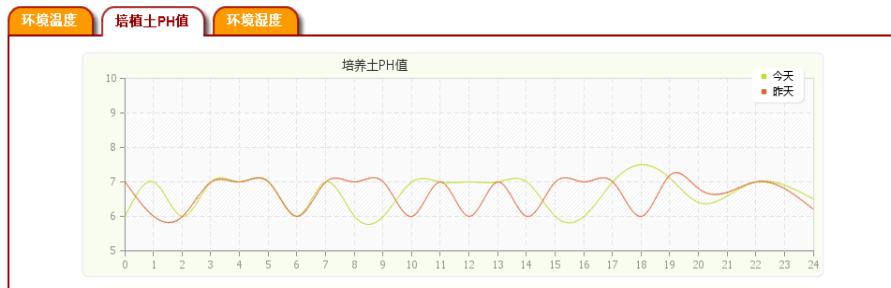


Fig. 4. Cultivated soil PH value data

Temperature monitoring diagram is shown in Fig.5.

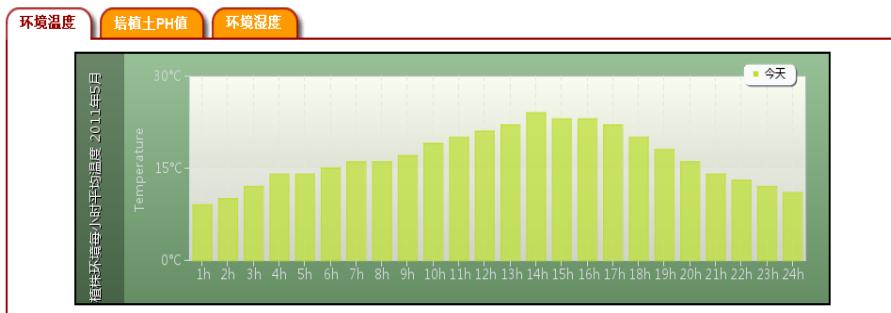


Fig. 5. Environment temperature data

Environment humidity monitoring diagram is shown in Fig.6.

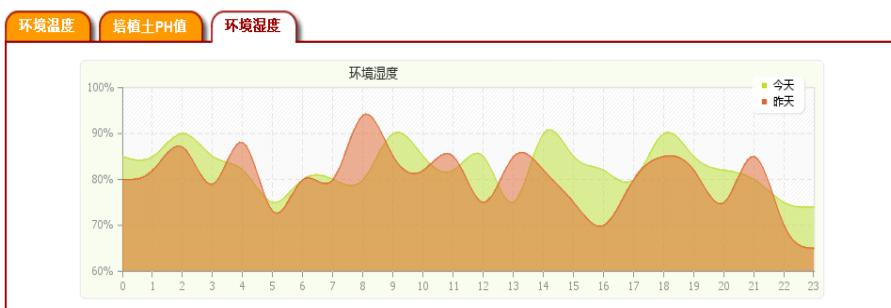


Fig. 6. Humidity data

## 5.4 Background Management System Design

Since management module is a critical module, the design is very strict. Program need to check login first. Only two conditions are met before granting access to database. Moreover, there is a timer design in this module, administrators will be logout automatically without operating system in half an hour. Core code is shown in Fig.7.

The crucial thing of backstage management is operating database, such as check, modify, add and delete. The most important is database connection and access.

```
define(ALL_PS,"SoSo");
if($_POST[submit]){$Member = M("Member");
$username = str_replace(" ","",$_POST['username']);
$data=$Member->getByusername($username);
$us = is_array($data);
$ps = $us ? md5($_POST['password'].ALL_PS) == $data['password'] : FALSE;
if($ps){$_SESSION['uid'] = $data['uid'];
$_SESSION['user_shell'] = $data['shell'];
md5($data['username'].$data['password'].ALL_PS);
$_SESSION['times']=mktime();}
```

**Fig. 7.** Sign Code

## 6 Conclusion

Field information monitoring platform based on The Internet of things combines data collection and monitoring warning. The internet of thing, computer software techniques and database techniques are used in management information. This platform satisfies completely function, easy to use, pages clean and beautiful. All of the features satisfy base requirements. This platform is valuable to use since it can be used in agricultural research, farmland monitoring and farmland management. This platform has broad prospects, which will deeply affect traditional agricultures.

**Acknowledgments.** This research was supported by GuangDong Provincial Science and Technology Planning Project of China under grant 2010B020315028.

## References

1. Weiser, M.: The Computer for the 21st Century. *Scimer* (September 1991)
  2. Satyanarayanan, M.: Pervasive Computing: Vision and Challenges. *IEEE Personal Communications*, 10–17 (August 2001)
  3. Welling, L., Thomson, L.: PHP and MySQL Web Development, 4th edn. Addison-Wesley Professional (October 2008)
  4. McArthur, K.: Pro PHP: Patterns, Frameworks, Testing and More. Apress (April 2008)

# Entropy-DEA Evaluation of Agile Supply Chain Management Based on IOT

Xiaowen Wang<sup>1</sup>, Dihui Lai<sup>2</sup>, Jinsheng He<sup>1,\*</sup>, and Shu'en Wang<sup>1</sup>

<sup>1</sup> Faculty of Management and Economics, Tianjin University  
Tianjin, China

wxwen7@126.com

<sup>2</sup> Department of Management Engineering, Tianjin Institute of Urban Construction  
Tianjin, China

ldhredarmy@vip.sina.com

**Abstract.** Information technology brought organizations revolution of business process reengineering. ASCM is a core strategy of enterprises. How to apply IOT into ASCM is a hot problem in management. This paper constructed a general model of ASCM based on IOT, and put forward the evaluation method through entropy-DEA. To demonstrate the theoretical model and it's evaluation method, we took a case as empirical analysis, and gave some advices from the conclusions.

**Keywords:** Internet of things, Agile Supply Chain management, Entropy-DEA Evaluation, Empirical Analysis.

## 1 Introduction

### 1.1 Internet of Thing

Internet of things (IOT) has become important development strategies in main counties and regions, for example "Intelligent Earth" of USA, "i2010" of EU, "U-Japan" & "i- Japan", "U-Korea" and "Feel China", etc since 2005. New forms of communication between people to thing, and between things themselves were enabled by embedding short-range mobile transceivers into a wide array of additional gadgets and everyday items, that is to say we will now have connectivity for anything from anytime, any place [1].

The technological basements of IOT are technologies of sensors, identity, spatial orientation, network communication, computation, management and servers such as radio frequency identification (RFID), infrared inductor, 2-dimensional bar code, laser scanners, Electronic Product Code (EPC)[2], global positioning system (GPS), IPv6, WiFi, UWB, ZigBee Bluetooth, Cloud computing, and special embedded software. Data can be acquired, stored, linked, transformed, transferred, analyzed, shared, and etc. through ubiquitous IOT, which can be widely implemented in many industries and areas for various benefits.

---

\* Corresponding author.

## 1.2 Research Significance

With the development of science, technology and productive forces, the need of customers for good quality, time, and specific has grown in buyer's market with intense competition since 1990s and some enterprises held supply chain management (SCM) as its core strategy to react the shorter product life cycle, the expanded types of products, and the higher exception to products and services.

Agility is the core property of SCM to cope with the rapid change of customers and environment outside, but asymmetric information is a general barrier troubled business. The efficiency of supply chain can be high improved with IOT by its excellent performance in information processing.

## 2 Agile Supply Chain Management Based on IOT

### 2.1 Agile Supply Chain and Agile Supply Chain Management

**Agile Supply Chain.** A direct supply chain is a system of three or more entities (organizations or individuals) directly involved in the upstream and downstream flows of products, services, finances, and information from one to another. An extended supply chain includes suppliers of the immediate supplier and customers of the immediate customer, and an ultimate supply chain includes all organizations involved in all the flows of products, services, finances and information from the ultimate supplier to the ultimate customer [3].

Agility is a business-wide capability that embraces organizational structures, information systems, logistics processes and, in particular, mindsets. A key characteristic of an agile organization is flexibility. Indeed the origins of agility as a business concept lies inflexible manufacturing systems (FMS) [4].

Agile supply chain (ASC) is a dynamic supply network with high flexibility and quick reactions which is integrated by suppliers, manufacturers, retailers and consumers through efficient integration and regulation of information, material and finance flow in competitive, cooperative and dynamic environment.

**Agile Supply Chain Management.** Agile supply chain management (ASCM) is the process of enterprise from procurement, production and distribution to prove flexibility in target of providing the right products/ services to right objects in the right time, right place and right quantity, ASCM gave much attention to time, information, and relationships.

**Systemic.** ASCM made some plan, coordination and constraints to flow of materials, information, and finance between partners in the supply chain system, to prove operation efficiency, security and make the whole benefit or utility maximization.

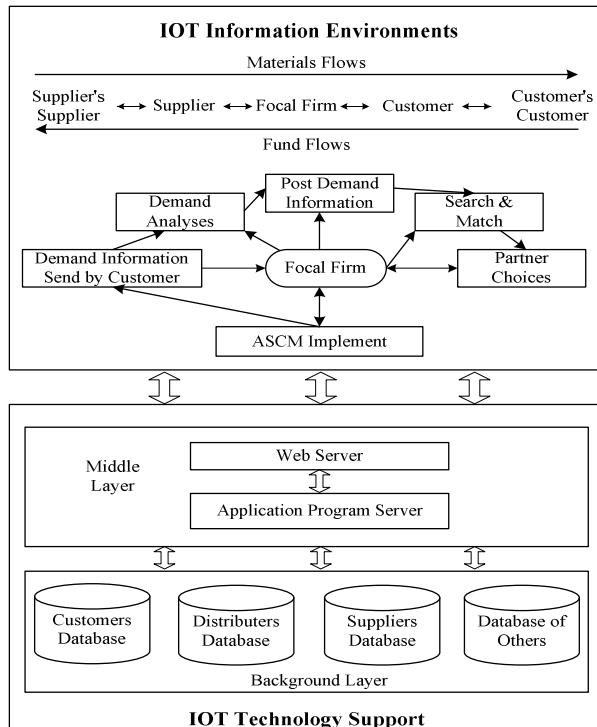
**Information Share.** Accurate and timely information can reduce uncertainty of logistics and finance in ASCM. Information integration plat based on information technology will give data share window and communication channel, to synchronize supply chain, drive production by order, and reduce stock.

**Agility.** Timeliness is one of key ingredients to improve supply chain efficiencies, and response in time is the core feature to feedback unpredictable market, and higher flexibility is a target of ASCM.

The same interest is the basement of ASCM participants to come together. Cooperation, empowerment, and sharing accountability between partners based on their credits, competitiveness, definite benefits and duties.

## 2.2 ASCM Model Based on IOT

Information technology brought organization revolution of business process reengineering, and the implements of IOT would made more data acquired, stored, transformed, transferred, mined, shared and the work of employees more easier, automatically, steadily, creative and valued.



**Fig. 1.** ASCM model based on IOT

Agile supply chain management model based on internet of things illustrated as following in Figure 1.

**Subjects.** The participants of ACM are supplier's supplier, supplier, focal firm, customer and customer's customer, materials flows from the first to the last, fund reversed, and information had double directions. The principle part of ASCM is focal firm, which is the most influential, either manufacturers, distributors, or other partners. The cooperation of subjects with enough information provided is apt to a set relationship of long and short, stable and dynamic, cooperative and competitive.

**Activities.** First, orders are main origins of most business nowadays, and demand information sent by customer to focal firm is the real beginning of ASCM. Second,

demand was analyzed to be accepted. Third, focal firm posted source demand information, chose appropriate partners after searching and matching. Then focal firm operated the agile supply chain based on IOT such as marketing sales, R&D, forecasting, production, purchasing, logistics, information system, finance and customer service, etc.

**IOT Information Environments.** The traits of IOT are so revolutionary, which pointed by anytime, anywhere, and anything, that all partners and activities of ASCM were in omnipresent environments of IOT information. The influences of IOT to demand, activities, manners, thoughts, cultures, etc. of organizations or individuals were far-reaching and definitive.

**IOT Technology Support.** IOT technology is the necessary technological basement of ASCM, which has two more layers. The most basic layer is background layer, which is composed of databases set of data through data acquiring, transforming, transferring and storing etc. such as database of customers, distributors, suppliers, and others. The upper layer is middle layer composed of Web Server and application program servers, which is vital to information transferring and integrating.

**Targets.** The targets of ASCM based on IOT are to improve Customer Satisfaction, realize the value/Profitability achieve and keep competitive advantages of subjects, to reduce operation cost, business cost, cycle of ASC and increase the flexibility of subjects and efficiency of ASC through information seamless linked and integrated of nodes.

In conclusion, ASCM model based on IOT is a system set of participants from source to consumption, purposing to accelerate, improve and reduce cost products and services, in core of industry, by means of management, and in base of IOT, driven by individual requirements and technology develop especially information technology.

### 3 Entropy-DEA Evaluation of ASCM Model Based on IOT

#### 3.1 Entropy-DEA Comprehensive Evaluation

**Introduction of Entropy and DEA.** Entropy is a measure of disorder or unpredictability for system. In information theory, entropy is a measure of the uncertainty associated with a random variable [5].

Data Envelopment Analysis (DEA) is an extreme point statistical method characterized as comparing each producer with only the "best" producers to evaluate the efficiency of a number of producers [6].

**Entropy-DEA Evaluation Model.** Nonlinear program in entropy-DEA illustrated as following in equation (1).

$$\begin{aligned}
 \min \quad & \omega^T X - \mu^T Y + \frac{1}{k} \left( \sum_{i=1}^m \omega_i \ln \omega_i + \sum_{r=1}^s \mu_r \ln \mu_r \right) \\
 (NP) \quad s.t. \quad & \frac{U^T Y_j}{V^T X_j} \leq 1, \quad j_0 = 1, 2, \dots, n \\
 & (X, Y) \in T \\
 & \omega_i > 0, \quad \mu_r > 0 \\
 & i = 1, \dots, m, \quad r = 1, \dots, s
 \end{aligned} \tag{1}$$

Define unpredictability Index H(X, Y):

$$H(X, Y) \triangleq \sum_{i=1}^m e^{-kx_i} + \sum_{r=1}^s e^{ky_r} \quad (2)$$

**Chaos Optimization Arithmetic for the Model.** Chaos optimization arithmetic has good performance to resolve nonlinear global optimization problems, which is suitable for random user equilibrium based on probability [7].

First, Logistic mapping was regarded as chaos carrier in equation (3).

$$z(n+1) = \mu \cdot z(n) \cdot (1 - z(n)) \quad (3)$$

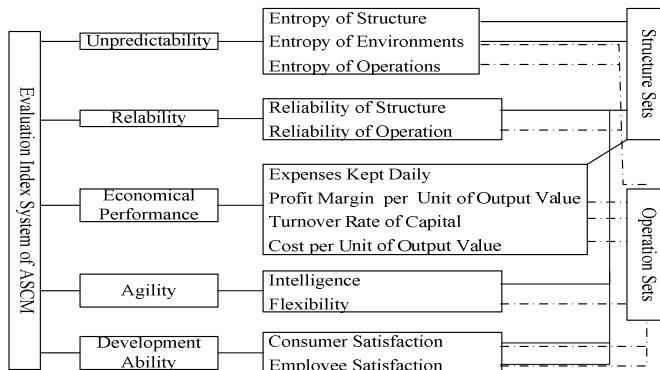
$z(n)$  above is the target function of nonlinear program equation (1), and  $z(n) \in [0, 1]$ ,  $\mu=4$ .

Second, optimized target function value was searching through steps initialization, adjusting the range of chaos variables after putting them into the model with carrier methods, searching repeated the minimum of target function by carrier irritation of chaos variables.

Last, searching would be finished if the criterions needed, and output the optimized values of variables and target.

### 3.2 Evaluation Index System

ASCM is a complicated system, in which the key ingredients may include four series of unpredictability, reliability, economic performances, and development ability in structure sets and operation sets respectively. The evaluation index system was illustrated following in figure 2.



**Fig. 2.** ASCM Evaluation Index System based on IOT

Structure sets refluxed system characters of relationships and significances between functions of ingredients in oppositely stainable and static situation of agile supply chain structure, which included indexes such as entropy of structure, entropy

of environment, reliability of structure, expenses kept daily, intelligence, consumer satisfaction and employee satisfaction. On the other hand, operation sets expressed evaluations of operations and randomness of ASCM operation process, which represented property of dynamic and innovation. In deed, entropy of environment is a factor external, which was set to unpredictability or structure set for convenience.

## 4 Empirical Analysis

Evaluation of ASCM of product S produced by group C were implemented with entropy-DEA model, the conclusions were commented and demonstrated, and then some advices were given.

### 4.1 Introduction of the Researched Manufacturer

Group C is an international advanced industrial enterprise with high level IOT implement all over works of the organization, excellence and innovation are their culture, competitive and cooperative relations linked objects of the organization inside and outside.

Company ltd. M is one of the wholly-owned subsidiaries, which produce manufacture S specially.

### 4.2 Data Collection

Data of the index preformed of product S from Jan to Dec in 2010 were collected from experts or functionaries. 5 input indexes are entropy of structure, entropy of environment, entropy of operation, expenses kept daily, cost per unit of output value, which will be littler, be better. 8 indexes remained are output ones.

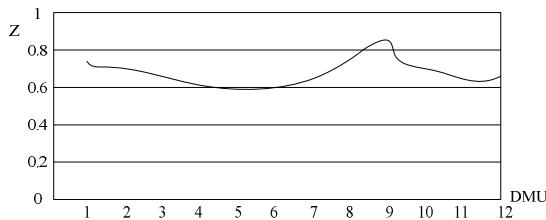
### 4.3 Computing Process

First, data collected initialized put into equation (1) as decision making unit denoted DMU1 for example, and object function were listed.

Second, questions were solved with chaos optimization arithmetic by Matlab program.

The outputs of program show that the minimal target function value is 0.724, after twice carries of chaos optimization, and then the optimized resolutions is  $\omega=[0.1753, 0.1461, 0.1065, 0.1472, 0.1588]$ ,  $\mu=[0.0179, 0.0122, 0.0164, 0.0540, 0.0219, 0.0159, 0.0160, 0.0023]$ .

Other optimized target functions of the remained nine evaluated ingredients were got through the same methods above. The consequences were listed following as  $DMU=[0.724, 0.698, 0.627, 0.614, 0.596, 0.605, 0.651, 0.764, 0.851, 0.707, 0.651, 0.660]$  in sequence, which presented by line in figure3.



**Fig. 3.** Value of optimized target functions

#### 4.4 Conclusions Illumination

**The Change of Fluctuations of Optimized Target Value of DMU.** The ASCM evaluation results for product S of company M were better in the end of 2010 than in the beginning or rather than from April to July, and the main reason of which is the corresponding orders. Intelligence based on IOT is bound to a great deal of repeated works, and when the orders were low down to some level, the cooperative effect would be counteracted. The business of product S had a certain extend of respectable seasonable.

**The Change of Ingredients of Evaluation Unit.** Entropy of structure and entropy of environments reduced month by month, which illustrated the ASCM system became more and more ordered by the learn and experience effect, and consistent with the ascending market share of product S of company M as the focal firm.

#### 4.5 Advices

**Non-core Functions Outsourcing.** The business of the researched object behaved seasonable, which brought some risk in increasing cost of marketing, human resources, and finance, etc. Non-core functions like logistics outsourced might be a constructive countermeasure to this problem, and could gave more resources and attentions to core functions to acquire and keep continually competitive advantages.

**Organization Learning.** The cooperative effect was promoting rapidly, rather than over loaded. Knowledge level updated and upgraded was now brought high economic performance, which and be made by special training, knowledge communications, and open thoughts of faults.

### 5 Conclusions

The technology of IOT brought revolutionary influence to business and our life, which made intelligence immanence. Agile supply chain management was core activity of enterprises in rapidly mutative market and technology environments, and the implements of IOT improved agility and core competitive advantages, which were systemic, dynamic, and undividable, in the ascending spire cycle of implement, evaluation, feedback, mending and improving.

**Acknowledgment.** Sponsor: National Natural Science Foundation of China NSFC70972116.

## References

1. International Telecommunication Union: ITU Internet Report 2005: The Internet of Things ITU, Geneva, p. 2 (November 2005)
2. Botthof, A., Bovenschulte, M., Evdokimov, S., et al.: The Internet of Things, pp. 16–17. Federal Ministry of Economics and Technology Pub., Berlin (2009)
3. Mentzer, J., DeWitt, W., Keebler, J., et al.: Defining Supply Chain Management. *Journal of Business Logistics* 22(2), 3–5 (2001)
4. Christopher, M.: The Agile Supply Chain Competing in Volatile Markets. *Industrial Marketing Management* 29(1), 37–44 (2000)
5. Shannon, C.: A Mathematical Theory of Communication. *Bell System Technical Journal* 27, 379–423 (1948)
6. Seiford, L., Thrall, R.: Recent developments in DEA: The mathematical programming approach to frontier analysis. *Journal of Econometrics* 46(1-2), 7–38 (1990)
7. Mu, D.: Research of Complexity and Evaluation Method of Supply Chain System, p. 192. Tsinghua University Publication, Beijing (2010) (in Chinese)

# AST-Based Plagiarism Detection Method

Liping Zhang, Dongsheng Liu, Yanchen Li, and Mei Zhong

Computer and Information Engineering College Inner Mongolia Normal University  
Hohhot, China  
cieczlp@163.com, cieclds@imnu.edu.cn

**Abstract.** In this paper, a code plagiarism detection based on the AST is studied. It pre-formats code, analysis lexical and syntax and obtains the corresponding AST. Then it traverses AST to generate code sequences, calculates the similarity of the code sequence and gets the code plagiarism detection report. Test results verify the effectiveness of the method.

**Keywords:** AST, plagiarism detection, ANTLR, similarity.

## 1 Introduction

Programming is an essential practice for colleges computer education, however, copying the code (or code plagiarism, copy the code) has also become a common cheating in programming course. Educational institutions abroad survey showed that 85.4% of the students admit plagiarism in homework [1-2]. The plagiarism is more and more serious in the country, influencing the quality of students and damaging the normal order of teaching. At the same time, many teachers have to repeat checking duplicate procedure on the general teaching and test, not only greatly consumed time and energy, but also can not ensure the accuracy and objectivity. Therefore, in addition to raise awareness education for students, the study of applicability and effective identification technology has become essential.

This paper mainly studies the code copy detection method based on abstract syntax tree (for short AST). First, preprocessing source code and formatting; then converting the code to AST; finally calculating similarity for the AST, thus obtaining the code plagiarism detection report.

## 2 Related Plagiarism Detection Methods

Early in the 1970s, foreign scholars began to research the technology and software to prevent large-scale copying program. There are two categories of commonly used code plagiarism detection tools: methods based on Attribute Counting (AC) and Structure Metrics (SM).

AC-based methods mainly consider the various statistical properties of the code; consider the structure of code very little. Halstead's Software Science metrics is the earliest and most typical attribute counting method [3].

SM-based methods judge the similarities mainly by analyzing internal structure of the code. General detection mechanism mainly includes two key steps: one is program standardization, in other words, the source into token string. Another is the similarity calculation, in other words, using specific technology to achieve pair wise comparison of program standardized output, calculating the similarity. At present, string matching technology is mainly used to realize the similarity calculation, the main methods are: Dotplot map [4], Levenshtein distance method, Longest Common Subsequence [5] and Looking for the longest public substring RKR – GST (Running-Karp-Rabin Greedy-String-Tiling) method [8-9].

Nowadays the code copy detection systems mostly use the method combining the means of AC and SM. Such as Moss system [8] of Stanford university in American, JPlag of Karlsruhe university in Germany [9], Sim of State university of Wisconsin [10], YAP3 [6] and SID [11] of the university of Sydney, etc. YAP3 and JPlag use RKR - GST algorithm, Sim uses a string array algorithm testing DNA sequence comparability and SID calculates similarity among the string using concept in information distance theory. These systems return a value between 0 and 1. But in the actual operation process, the teacher must step in to check many students program and then chooses a suitable threshold. Research has shown that when using JPlag and Moss, code similarity will obviously decreased when adding some redundant statements and statement redundant variables or split statements in the student program.

Domestic studies are relatively few in this field. Our team started to research the code copy detection technology in early 2002, from two aspects: source and non-source code simultaneously researched code plagiarism detection, recently obtained the certain results [12-15].

### 3 Code Plagiarism Detection Based on the AST

Programming is an essential practice for colleges computer education, however, copying the code (or code plagiarism, copy the code) has also become a common cheating in programming course. Educational institutions abroad survey showed that 85.4% of the students admit plagiarism in homework [1-2]. The plagiarism is more and more serious in the country, influencing the quality of students and damaging the normal order of teaching. At the same time, many teachers have to repeat checking duplicate procedure on the general teaching and test, not only greatly consumed time and energy, but also can not ensure the accuracy and objectivity. Therefore, in addition to raise awareness education for students, the study of applicability and effective identification technology has become essential.

This paper mainly studies the code copy detection method based on abstract syntax tree(for short AST). First, preprocessing source code and formatting; then converting the code to AST; finally calculating similarity for the AST, thus obtaining the code plagiarism detection report.

### 3.1 Pretreatment

In order to detect plagiarism of program code more effectively, improve the testing efficiency, we make a series of pretreatment and formatting source code before converting the code. For instance, in order to prevent the plagiarism adding annotation, empty line or extra space, we remove all the annotation, empty line and extra space in program; for the same programming, the program header files are roughly same, in order to prevent the plagiarism joining the useless header files in copying, all the header files are removed from program. Formatting pretreatment includes the following aspects:

- a) Removing all annotation, empty line in program.
- b) Removing header files when it does not affect program semantic.
- c) Adding a pair of braces for the control structure containing a statement when it does not affect program semantic and there are no paired braces.

### 3.2 Code Conversion

Code conversion is converting the code after preprocessing to the corresponding AST through lexical analysis and syntax analysis to provide standard mathematical models for plagiarism detection.

#### 1) Formal Form of the Code

After the pretreatment, the source code are converted to a form of a higher abstraction, such as token, string, AST, PDG, eigenvector etc. In this paper we use AST to express the source code. Therefore, transforming the source code to AST provides good mathematical model of the standard framework for the follow copy detection.

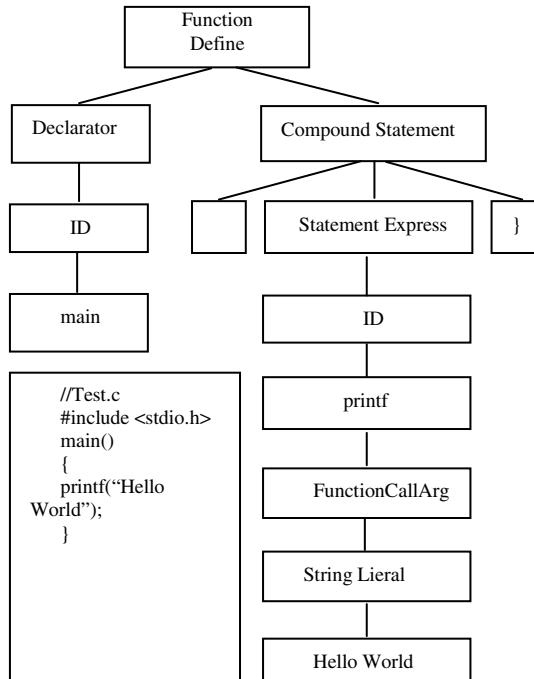
#### 2) Using ANTLR to construct AST

ANTLR (Another Tool for Language Recognition) is an open-source grammatical analysis tool, ANTLR is powerful, extending easily comparing with the lex/Yacc traditional grammar analytic tool. And it uses top-down analysis program, easy for reading comprehension. In grammar analytical process, choosing ANTLR construct AST is because of two reason, one is that it provides good support in creating grammar trees, traversing and transforming trees; the other is that as a language analysis tools, as long as offering grammar files, it can analyze source codes and deal with multilingual source code of Java, C and C++ etc, has very good generality and expansibility.

Before using ANTLR, the user should formulate grammar files; write the custom lexical analyzer and grammar analyzer rules in. ANTLR produces codes corresponding lexical analyzer and grammatical analyzer by reading the rules in analysis grammar document. Using the generated lexical analyzer, first we convert the input code to the flow composed by phrases, and then use this flow as input of lexical analyzer to get the final result –AST. Figure 1 shows the example that the source code Test.c is analyzed by ANTLR to get AST.

### 3) Formulating Grammar File

The key point we study in this article is the similarity of code. So in order to avoid the heavy work of formulating we use grammar file published on the ANTLR official website and make some necessary extensions and modifications. Using C language grammar files as an example we focus on two areas:



**Fig. 1.** AST form of the source cod segment

- a) The grammar file published on the official website only supports C89 standard. For example: the definition of the variable can only be in the beginning of the variable declaration part of the function, no supporting loop variable definition in scope of “for” sentence and so on. In order to make it support most of the C99 standards, in this paper, we have done the necessary expansion to the grammar file.
- b) In order to meet the similarity calculation method, we fix the length of the virtual token to maximum reduce the length of the output string sequence, thus the efficiency of the matching calculation is improved.

### 3.3 Plagiarism Detection

Copy detection is divided into two steps: first to traverse the AST and generates code sequences; then to calculate similarity of code sequence, finally get the final copy detection report.

### 3.3.1 Traverse the AST

Arithmetic generating code sequences by traversing AST is showed by figure 2. The arithmetic is start from the root node of AST, makes the depth-first traversal, collects only the internal AST node and skips the leaf node in the process of traversing. Because leaf nodes contain the specific code information, such as identifiers, variable names, literal, etc. The information is interference information for the plagiarism detection and should not be used in the detection process. In addition, AST contains sign information (token) in each node; we can get node location in source program from the sign. So even the internal nodes can be converted to the original code, thus marks the position of the similar code in the final copy detection report.

```

import java.io.*;
import ANTLR.CommonAST;
import ANTLR.collections.AST;
class Calc
{
    public static void main(String[] args) {
        CalcLexer lexer = new CalcLexer(new
            DataInputStream(System.in));
        //input string
        CalcParser parser = new CalcParser(lexer); //create parser
        Parser.expr(); // analysis the expressions
        //generate syntax tree
        CommonAST t = (CommonAST)parser.getAST();
        //output tree with LISP tags
        System.out.println(t.toStringList());
        //generate tree with parser
        CalcTreeWalker walker = new CalcTreeWalker();
        String r = walker.expr(t); // retrieve the tree
        System.out.println(r);
    }
}

```

**Fig. 2.** Arithmetic generating code sequences by traversing AST

### 3.3.2 GST Algorithms

The similarity of code sequences uses Greedy String Tiling algorithms (GST) to find the largest public substring of two codes. Some famous student program plagiarism detection system such as Jplag, adopted this algorithm. Pseudo codes of the algorithm are shown in Figure 3.

GST algorithm need to find all maximum matches as possible through repeated testing, therefore, the time complexity of the algorithm is relatively low. Worst case time complexity is  $O(n^3)$ , best case time complexity is  $O(n^2)$ . In this paper we further optimized GST arithmetic. For worst case, the time complexity is still  $O(n^3)$ . However it is usually less than  $O(n^2)$  in practice. Optimization includes the following two aspects.

- 1) GST algorithm's time complexity is clearly proportional to the length of the string. Therefore shortening the length of the string sequence can effectively improve the

efficiency of the matching. Token of each node is fixed of two characters length, thus the length of the character sequence is shortened as short as possible. Figure 4 is an example. The token in AST code sequence node is often long for the readability and the length is not fixed. This is shown in Figure 5. Fixed length token is shown in Figure 6.

- 2) Because the token length is fixed and semantics of each token are fixed, when we calculate the hash value of each substring, we can skip one character each time. The number of hash value (in String A for example) is reduced from  $|A| - s + 1$  to  $(|A| - s)/2 + 1$ , thereby reducing the number of comparisons and improving the efficiency of the algorithm.

```

Greedy-String-Tiling(String A ,String B){
    Tiles={ }
    Do{
        maxmatch=MinimumMatchLength;
        matches ={ };
        Forall unmarked tokens Aa in A{
            Forall unmarked tokens Bb in B{
                j=0;
                while(Aa+j==Bb+j)&&
                    unmarked(Aa+j)&&unmarked(b+j);
                j++;
                if(j==maxmatch)
                    matches=matches ⊕ match(a,b,j);
                else if(j>maxmatch){
                    matches={ match(a,b,j)};
                    maxmatch=j;
                }
            }
        }
        Forallmatch(a,b,maxmatch)∈ matches{
            For j=0... (maxmatch-1){
                mark(Aa+j);
                mark(Bb+j);
            }
            Tiles =Tiles ∪ match(a,b,maxmatch);
        }
    }while(maxmatch > MinimumMatchLength);
    retrun Tiles
}

```

**Fig. 3.** GST algorithms

```

for(i=0;i<n;i++)  sum+=aa[i];
sum=sum/n;

```

**Fig. 4.** Fragment of C code

```

LITERAL_for ASSIGNID Number LT ID ID NpostfixExp ID INC NstatementExpr
PLUS_ASSIGN ID NpostfixExp ID LBRACKET ID RBRACKET NstatementExpr
ASSIGN ID DIV ID ID

```

**Fig. 5.** Token sequence of AST

NS	xh	AS	ID	UM	LT	ID	ID	NP	ID	IN	SX	PA	ID	NP	ID	LB	ID	RB	SX	AS	ID	DC	ID	ID
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

**Fig. 6.** AST sequence of fixed length of token

### 3.3.3 Calculating the Similarity of the Codes

In this paper the similarity of the code between A and B is calculated according to the following formula. Using GST arithmetic we can get maximal matching set tiles of code A and B. And using this set we can calculate the similarity of AST sequence of two programs. Using the results the similarity of two corresponding program code A and B can be find.

$$sim(A, B) = \frac{2 \cdot coverage(tiles)}{|A| + |B|} \quad (1)$$

$$coverage(Tiles) = \sum_{match(a,b,length) \in Tiles} length \quad (2)$$

## 4 Experiments and Analysis

Materials: for the binary search problem, construct a testing set containing nine C language programs, each source contains an average of about 60 lines of code. 1.cpp and 2.cpp in the test set are the original program completed independently, 3.cpp to 9.cpp are modified from 2.cpp and in the modified process, we have imitated plagiarism methods: complete copy, rename an identifier, a single statement or block of code rearrangement, add and delete notes, change the control logic and so on. Experimental results statistics below. Comparison of run-time efficiency is as follows.

Before improvement:

```
7.cpp.ast<>9.cpp.ast: 0.529412
8.cpp.ast<>9.cpp.ast: 0.965517
The time was: 0.562000
```

After improvement:

```
7.cpp.ast<>9.cpp.ast: 0.529412
8.cpp.ast<>9.cpp.ast: 0.965517
The time was: 0.156000
```

From the figures above we can see that efficiency has been greatly enhanced after the GST algorithm improvements.

## 5 Summary

This paper proposes a code copy detection method based on AST. This method has been achieved and integrated into the C code plagiarism detection experiment system. Analysis and test show that the method could realize the program code plagiarism detection, achieves the expected effect basically, and also has deficiencies. Similarity will low to some extend when using the method to analyze expression splitting and replacing the control logic plagiarisms.

Next we will make up deficiencies for the exist, and continue to research: ① Expanding the scale of experimental system test set, collecting additional procedures to verify the method; ②The existing copy detection do not include clustering analysis

and unable to get "copying gang", we should make clustering analysis ; ③Because of the time, the experimental system achieves C code plagiarism detection. For a variety of programming languages, we should develop and refine the grammar files of C++, Java and other programming languages to realize the copy detection of various languages, so as to improve the commonality and expansibility of experiment system.

**Acknowledgments.** This work is supported by the National Natural Science Foundation of China under Grant No. 60940027 and the National Natural Science Foundation of Inner Mongolia under Grant No.2010MS0906.

## References

1. Georgina, C., Mike, J.: Source2code plagiarism: A UK academic perspective. Research Report RR-422, Department of Computer Science, University of Warwick (2006)
2. Sheard, J., Dick, M., Markham, S., et al.: Cheating and Plagiarism: perceptions and practices of first year it students. In: Proc. of the 7th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education, pp. 183–187. Association for Computing Machinery, New York (2002)
3. Halstead, M.H.: Elements of Software Science. Elsevier Science Inc., New York (1977)
4. Granville, A.: Detecting Plagiarism in Java Code, Wilks, Y.(Supervisor) (2002)
5. Clough, P.: Plagiarism in Natural and Programming Languages: an Overview of Current Tools and Technologies. Research Memoranda CS-00-05, Department of Computer Science, University of Sheffield (2000)
6. Wise, M.J.: YAP3: Improvement Detection of Similarities in Computer Program and Other Texts. In: Proc. of the 27th SIGCSE Technical Symposium on Computer Science Education, vol. 28(1), pp. 130–134. Association for Computing Machinery, New York (1996)
7. Prechelt, L., Malpohl, G., Philipsen, M.: Finding Plagiarisms among a Set of Programs with JPlag. Journal of Universal Computer Science 8(11), 1016–1038 (2002)
8. Aiken, A.: Moss: a system for detecting software plagiarism (EB/OL) (2006), <http://theory.stanford.edu/~aiken/moss/> (February 01, 2009)
9. Emeric, K., Moritz, K.: JPlag: a system that finds similarities among multiple sets of source code files (EB/OL) (2005), <http://www.ipd.Uni-karlsruhe.de/jplag/> (February 01, 2009)
10. Gitchell, D., Sim, T.N.: A utility for detecting similarity in computer programs. In: Proc. of the 30th SIGCSE Technical Symposium on Computer Science Education, pp. 266–270. Association for Computing Machinery, New York (1999)
11. Chen, X., Francia, B., Li, M., et al.: Shared Information and Program Plagiarism Detection. IEEE Transaction on Information Theory 50(7), 1545–1551 (2003)
12. Zhang, L., Liu, D., Li, Y.: Copy detection and evaluation mechanisms based syntax tree code. Inner Mongolia University 41(5), 594–600 (2010)
13. Zhong, M., Zhang, L., Liu, D.: XML-based plagiarism detection algorithm C code. Computer Engineering and Applications, 215–218 (2010)
14. Zhong, M., Liu, D.: Plagiarism Detection Model for C Program. In: The 3rd International Conference on Advanced Computer Theory and Engineering, Chengdu, pp. 460–464 (2010)
15. Li, Y., Liu, D.: Suffix Tree Based Plagiarism Detection Method for C Code. In: 2010 International Conference on Future Computer, Control and Communication, Nanning, pp. 210–213 (2010)

# A Research on Plagiarism Detecting Method Based on XML Similarity and Clustering

Shengying Jia, Dongsheng Liu,  
Liping Zhang, and Chenglong Liu

Computer and Information Engineering College Inner Mongolia Normal University  
Hohhot, China  
{jsy\_1985, cieczlp, liu\_chengl}@163.com,  
cieclds@imnu.edu.cn

**Abstract.** This paper mainly studies on plagiarism detection method based on the XML similarity and the clustering. Firstly, it analysis the replication types and the copy features of the programming source code; Secondly, extracts the specific code strings and the line positions, and then converts it into the XML text; Thirdly, generates a visualization detecting report according to the procedure similarity; Finally, we use the clustering method to identify the plagiarism “cluster” and the “source” of the plagiarism.

**Keywords:** XML, Plagiarism detection, Clustering.

## 1 Introduction

Nowadays, many teachers who in the most colleges and universities use computer systems for teaching and assessment have become a common method. Particularly because of the strongly engineering practicality, the programming courses almost exclusively dependents on computers for teaching and assessment. In order to facilitate, many students often use the other students' program as a template, modifying it briefly, or even handing teacher with unchanged program which attempt to muddle through, this is a very irresponsible act to them. Particularly in the information age, we search resources more convenient, which leads to the phenomenon of plagiarism are more serious. Therefore, suppressing plagiarism is needed urgently. In addition, it is a very complicated work for teachers to determine the plagiarism. Traditional artificial methods which can determine the plagiarism was both a waste of time and effort, moreover, the final detecting result has a great relationship between tester's experience and the program scale. For example, giving two or more program code, it is very difficult to detect whether they are same or similar. Supposing there are n programming codes to determine, it requires to matching  $n*(n-1) / 2$  times, when the number of procedures are large quantities, this will be a very time-consuming and labor-intensive work.

## 2 Relevant Research

Early as the 1970s, some foreign scholars began to study technology and software for detecting program code plagiarism. So far, there are two popular used code plagiarism detection technologies: Attribute Counting (AC) and Structure Metrics (SM).

### 2.1 Attribute Counting

Attribute counting approach calculates the variety of the property information in the program, however, do not consider the structural information in the program. Halstead's Software Science metrics is the earliest and the most typical attribute counting approach [1-2]. In this approach, they counted the following four values for each program:

- $\eta_1$ = the number of single operator ;
- $\eta_2$ = the number of single operand ;
- $N_1$ = the total number of all operators;
- $N_2$ = the total number of all operations.

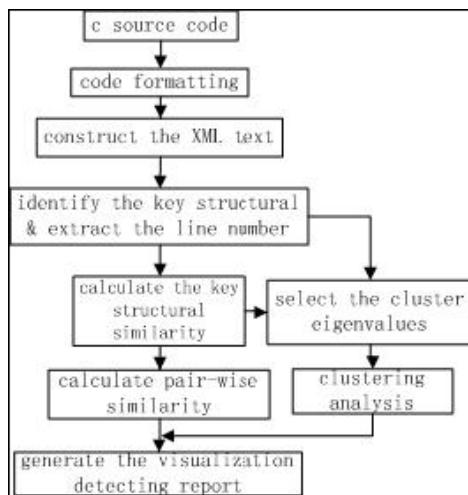
According to these four basic properties, definition:  $\eta=\eta_1+\eta_2$  is the vocabulary,  $N=N_1+N_2$  is the length of implementation, next calculated the program's capacity  $V = N \log_2(\eta)$ . And then combined this information into a characteristic vector  $H(\eta, N, V)$ . After generated the characteristic vector for every program which to be detected, next calculated the distance between each of the two vectors, if the distance between the two programs vector is very small, we can think that these two programs are similar. Later, Faidhi and Robinson developed a somewhat more complicated plagiarism detecting system [3], they increased some functions which compares the quantity of two procedures for controlling structure in the system. Since then, there are many attribute counting plagiarism detecting system published, for example, Verco KL and Wise MJ [4] developed an automated plagiarism detecting systems, etc.

### 2.2 Structure Metrics

Structure metrics approach detects the similarity mainly through analysis and compares to the internal structure of the program. General detection mechanism includes two key steps: One is standardized the procedures, another is the similarity calculation. Some famous systems are: Stanford University developed the system of Moss [5] (Measure of Software Similarity), University of Karlsruhe, Germany developed JPlag [6-7], Wichita State University developed the system of Sim [8], University of Sydney has developed a series of YAP [9], Osaka University has developed the system of CCFinder [10], U.S. Semantic Designs company has developed CloneDR [11-12] and CP-Miner [13] and so on.

### 3 Design and Implementation of the Experimental System

This paper focuses on an approach for code plagiarism detection which combines the XML similarity and clustering. Firstly, proposing a method for extracts the program specific code strings, using XML text to store the key structure information of the program which vulnerable to plagiarism. Secondly, according to the different key structure, the different algorithm of generating an XML text from program code is designed, and the similarity of two programs is calculated by the line similarity of XML text. And then, we use the similarity of programs to execute the clustering analysis. Finally, uses the above result to generate a visualization detecting report. The implementation procedure is shown in Figure 1.



**Fig. 1.** System Framework

#### 3.1 Construction of XML Text

Among the procedural programming languages, the expression of the same logical statement often varied. It usually not a simple copy of a program code when plagiarism occurs, but rather has various forms of transformation. 2001 Edward L. Jones [14] summarized ten categories of program plagiarism means. Based on this, in 2008 Beijing Aerospace University Zhao Changhai [15] added two plagiarism means, including the constant replacement and the expression split. Such as: in the C source code, the select structure can use both if statement and switch statement, the loop structure also have three kinds of expression: for, while and do-while, and the three expressions can be interchangeable under certain conditions. This paper summarizes the plagiarism means of the different key structures, and then determines which key information need to be extracted to convert to XML text, the information consists of two parts: some including parts of program code which can represent the key structure

[16], others including the tag information which representatives start and end position of the key structure. Finally, according to the different key structure and corresponding means of the plagiarism, the different algorithm of generating an XML text from program code is designed. The transformed XML text is shown in Figure 2.

<pre> int main() { int i,sum=0,num; for(i=0;i&lt;10;i=i+1) { sum=sum+num; } </pre>	<pre> &lt;code&gt; &lt;function pos="1" name="main" returntype="int" number="0"&gt; body={int i,sum=0,num for(i=0;i&lt;10;i=i+1){sum=sum+num,}}&lt;/function&gt; &lt;variables pos="3"&gt;&lt;int i,sum=0,num/&gt; &lt;var type="int" name="i" value="0" big=""/&gt;&lt;/var&gt; &lt;var type="int" name="sum" value="0" big=""/&gt;&lt;/var&gt; &lt;var type="int" name="num" value="0" big=""/&gt;&lt;/var&gt; &lt;/variables&gt; &lt;other&gt;pos="3" sum=0,&lt;/other&gt; &lt;other&gt;pos="4" i=0,&lt;/other&gt; &lt;for pos="4" conditions="i&lt;10" body={(sum=sum+num,)}&gt; &lt;other&gt;pos="6" sum=sum+num,&lt;/other&gt; &lt;other&gt;pos="4" i=i+1,&lt;/other&gt; &lt;/for&gt; &lt;/function&gt; &lt;/code&gt; </pre>
--	--

**Fig. 2.** The left figure is the code after formatting and the right one is the code converted to the XML text

### 3.2 Similarity Calculation

In this part, according to the different key structure and the corresponding means of plagiarism, we design the different similarity comparison algorithm. First, identify the key structure what this line statement for. Then call for different algorithm to calculate the similarity value of the XML text line. Finally, according to the similarity of XML text line, calculate the pair-wise similarity [16]. Similarity is calculated as follows:

$$\text{Sim}(A, B) = \frac{\text{sum\_row}(A) + \text{sum\_row}(B)}{|A| + |B|} \quad (1)$$

$$\text{sum\_row}(A) = \sum_{i=0}^{|A|-1} \text{simi}(i) \quad (2)$$

$$\text{sum\_row}(B) = \sum_{i=0}^{|B|-1} \text{simi}(i) \quad (3)$$

$\begin{cases} |A|, |B|: \text{representing total number of procedures A and} \\ \text{B corresponding the XML text.} \\ \text{simi}(i): \text{representing similarity of i-line of the XML text.} \end{cases}$

### 3.3 Clustering Analysis

Most of the studies on the similarity of the source code only stay in the calculation the similarity of the pair-wise program. It is rarely concerned about the further comparative analysis of the results which has been calculated. In this paper, based on

the results of similarity calculation, we use the clustering analysis in the data mining, combine with the program's key structure information which has extracted from the last phase, generate the corresponding eigenvalue and finally find the copy of the "cluster."

Clustering is segments the data into different classes or clusters which according to certain standard, so the similarity between data objects in the same cluster as large as possible, while the data objects in the different cluster are also different as large as possible. We suppose each data object  $x_i$  has multiple attributes (or eigenvalue)  $x_i = (x_{i1}, x_{i2}, \dots, x_{id})$ , each property  $x_{ij}$  can be both numeric and enumerate type. This paper proposes a clustering approach which can carry out the suspicious plagiarism programs. Using the detecting results which last phase obtained. Selecting the key structural as the eigenvalue, and then quantitative it, so that all the eigenvalue of data objects can be seen as a matrix. The basic idea of the approach as follows:

- 1) Extracting the key structure which the similarity is larger than a certain threshold as the program's properties.
- 2) According to each properties' proportion in the program, they are quantified as a group vector which constituted by the keyword weight.
- 3) Comparing the unclassified program's similarity with every programs in a existing cluster. If the similarity is larger than a certain threshold, then incorporating this program into this cluster, while the similarity value is relatively smaller, then the program is classified as a new cluster.
- 4) Repeating the above steps until all the program compared completely.

Through above clustering analysis, all the being analyzed programs are divided into several clusters, each "cluster" represents a plagiarism "group." Next, according to the similarity which directly plagiarism is larger than indirectly plagiarism in each clusters, we can calculate the "source" plagiarism program of corresponding structure.

### **3.4 Generating the Visualization Detecting Report**

Based on the previous research, the final detecting results just given the similarity value between the pair-wise programs, the degree of the similarity is not reaction too much in the source code. Therefore, the plagiarism target cannot be effectively locked. Since the source code have the natural similarity, for example: if the program needs to solve some less difficult problems, corresponding, the process is simply, such a program would have large similarity. In addition, affects by the skills training and the mindset, different programmers may be having the same idea to solve the same problem, the code would be similar either. Or in the course of teaching, students who at the same learning environment or the same learning phase write the program code may be have certain similarities. Therefore, it couldn't reflect whether plagiarism or the degree of plagiarism just through the pair-wise similarities result.

In this paper, we summary of the above detecting information together, finally generate a visualization detecting report. The report including follows: the pair-wise

similarity, the plagiarism lines of code, the beginning and ending line number of the plagiarism section and their corresponding file names and other information. Figure 3 shows the parts of two program's visualization detecting report.

```

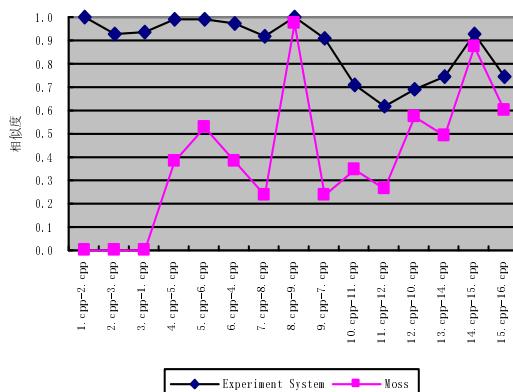
result.txt - 记事本
文件(F) 编辑(E) 格式(U) 查看(V) 帮助(H)
xml\1.cpp.xml----xml\2.cpp.xml: 41%
1.cpp(4-7)----2.cpp(5-6)
1.cpp(10-15)----2.cpp(8-12)
1.cpp(19-26)----2.cpp(22-27)

```

**Fig. 3.** Visualization Detecting Report

## 4 Experiment and Analysis

We selected 15 C language programs to test the similarity. These programs came from five different jobs which represented different plagiarism means. Detecting results compared with the Moss are shown in the Figure 4. From the experimental results we can see that the MOSS system has strong noise in detecting replace the control structures for the equivalent. In contrast, the approach which this paper proposed has better detecting results than the MOSS system. The effect is especially remarkable in the increasing the redundancy, replacing the control structures for the equivalent and renaming the identifier and so on.



**Fig. 4.** Detecting results complared with MOSS

Then we selected another 20 C language programs to test the clustering analysis. These programs include four “source” codes, each “source” codes have three plagiarism codes, and the rest programs are the code which has the same plagiarism means. The result of the clustering analysis is shown in the Table 1. From the sheet we can see that the clustering method proposed in this paper could successfully classify the plagiarism “group”.

**Table 1.** Clustering Results

Plagiarism "group"	Manual detecting	Clustering results
1	{1,5,11,15,18}	{1,5,11,12,13,15,18}
2	{2,6,12,14,17}	{2,7,14,17}
3	{3,7,8,10,19}	{3,8,13,19}
4	{4,9,13,16,20}	{4,9,10,16,20}

## 5 Conclusions

This paper introduces a plagiarism detection model and its implementation combined the XML similarity and the clustering. In the adaptive verification system, the experimental system can detect the following plagiarism means: complete copy, modify the comment, re-layout, identifier rename, code block reorder, reorder the statement in code block, change the operator or operand order of the expression, change the data type, replace the control structures for the equivalent, split the expression and increase the redundancy.

According to the existing plagiarism detecting systems, program's plagiarism detecting results did not respond too much on the source code, therefore, the target of the plagiarism cannot be effectively locked. This system not only can detect the common plagiarism means appears from the program, but also generate the visualization detecting report which can further analysis of the finally results, it can even display the plagiarism "group" classified, find out the "source" code in the plagiarism "group". Teachers can do the further analysis based on this detecting report.

Since this experiment is limited study, experiment ranges relatively narrow. Next work, we will continue to gather additional programs to verify the effectiveness of the system, moreover, we will apply the system to the programming courses and the program competition to detect the program which the student write.

**Acknowledgments.** This work is supported by the National Natural Science Foundation of China under Grant No. 60940027, the National Natural Science Foundation of Inner Mongolia under Grant No.2010MS0906 and the Graduate Scientific Research and Innovative Foundation of Inner Mongolia Normal University CXJJS11067.

## References

1. Halstead, M.H.: Elements of Software Science (1977) ISBN: 0444002057
2. Ottenstein, J.K.: An Algorithmic Approach to the Detection and Prevention of Plagiarism. ACM SIGCSE Bulletin 8(4), 30–41 (1976)
3. Faidhi, J.A.W., Robinson, S.K.: An Empirical Approach for Detecting Program Similarity and Plagiarism within a University Programming Environment. Computers and Education 11(1), 11–19 (1987)

4. Verco, K.L., Wise, M.J.: Plagiarism à la mode: A comparison of automated systems for detecting suspected plagiarism. *The Computer Journal* 39(9), 741–750 (1996)
5. Aiken, A.: Moss: a system for detecting software plagiarism (EB/OL),  
<http://theory.stanford.edu/~aiken/moss/2009-12-21>
6. Prechelt, L., Malpohl, G., Philippsen, M.: Finding Plagiarisms among a Set of Programs with JPlag. *Journal of Universal Computer Science* 8(11), 1016–1038 (2002)
7. Emeric, K., Moritz, K.: JPlag: a system that finds similarities among multiple sets of source code files (EB/OL),  
<http://www.ipd.Uni-karlsruhe.de/jplag/2009-12-21>
8. Gitchell, D., Sim, T.N.: A utility for detecting similarity in computer programs. In: Proc. of the 30th SIGCSE Technical Symposium on Computer Science Education, pp. 266–270. Association for Computing Machinery, New York (1999)
9. Wise, M.J.: YAP3: Improvement Detection of Similarities in Computer Program and Other Texts. In: Proc. of the 27th SIGCSE Technical Symposium on Computer Science Education, vol. 289(1), pp. 130–134. Association for Computing Machinery, New York (1996)
10. Kamiya, T., Kusumoto, S., Inoue, K.: CCFinder: a multilingualistic token-based code clone detection system for large scale source code. *TSE* 28(7), 654–670 (2002)
11. Baxter, I.D., Yahin, A., Moura, L., Sant'Anna, M., Bier, L.: Clone detection using abstract syntax trees. In: ICSM, pp. 368–377 (1998)
12. Wahler, V., Seipel, D., von Gudenberg, J.W., Fischer, G.: Clone detection in source code by frequent itemset techniques. In: SCAM, pp. 128–135 (2004)
13. Li, Z., Lu, S., Myagmar, S., Zhou, Y.: CP-Miner: A tool for finding copy-paste and related bugs in operating system code. In: OSDI, pp. 289–302 (2004)
14. Jones, E.L.: Metrics based plagiarism monitoring. In: Proceedings of the Sixth Annual CCSC Northeastern Conference on the Journal of Computing in Small Colleges, vol. 16(4), pp. 253–261 (2001)
15. Zhao, C., Yan, H., Jin, M.: Approach based on compiling optimization and disassembling to detect program similarity. *Journal of Beijing University of Aeronautics and Astronautics* 34(6) (2008)
16. Zhong, M., Liu, D.: An XML plagiarism detection model for C program. In: Proc. of 2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE), pp. 460–464 (2010)

# Research on the Relevant Standards of Internet of Tings

Jie Jing and Hao Li

Research Institute of Electronic Science and Technology,  
University of Electronic Science and Technology  
Chengdu, China  
jingjie33@yahoo.cn

**Abstract.** At present, the development of the relevant standards with independent intellectual property rights for the internet of things (IoT) will directly determine our IoT industry's development and national interests in the future. Based on the research of ISO/IEC 18000 series of standards, EPC standards and IEEE802.15.4 (Zigbee) protocol, this article puts forward several key points that should be focused on during the establishment of standards for IoT, proposes the suggestion for content of standards and technique routing for the establishment.

**Keywords:** IoT, related standards of IoT, the formulation of standards.

## 1 Introduction

After Massachusetts Institute of Technology's (MIT) Center for Automatic-Identification System (Auto-ID Center) proposed the concept of "Internet of Things" which was called as Electronic Product Code (EPC) system in 1998, "Internet of Things" is more and more concerned, it has been the third wave of information industry after the computer, Internet and mobile communication network in the word. In addition to the participation of computer networks, mobile communications networks and cloud computing, with the rapid development of informationization and intelligentize, IoT has a wide range of application and extensive requirement. IoT technology in China is still in the initial stage, and it's application is only in small-scale and closed-loop model, the most critical factor of this status is the establishment of standardization system, which constrained the application of IoT on a large scale. Lack of standards will restrict the technology's development and application, unified and open standards is the inevitable precondition for the promotion of IoT, meanwhile, establishment of standard determine the control rights and property rights of information, related to national development, information security and technology strategy in correlated industry.

## 2 The Investigation of Relevant Standards of IoT

The relevant standards of IoT around the world mainly include: ISO/IEC、EPC、UID、AIM、IP-X. These standards focus in different points of emphasis and do not reach a

unified. RFID and ubiquitous sensor network is the most widely used technology in Perceptive layer, the establishment of standard for these two technology is especially important. Currently, the dominant standard on the RFID is EPC standard and ISO/IEC 18000 standard, IEEE 802.15.4 protocol is recently proposed for ubiquitous sensor networks.

### (1) ISO/IEC 18000

ISO is the most important International Organization for standardization, it had set the ISO/IEC 18000 series of standards which is the earliest standard about IoT, the standard includes four aspects which are data encoding, air interface, testing and applicative standard. ISO/IEC 18000 standards work in five different bands of frequency (below 135KHz, 13.56MHz, 2400-2483.5MHz, 860-960MHz, 433.92MHz), Series of standards of RFID on the basis of basic standard made by ISO, according to different objects' specific requirements of particular application, have formulated working conditions, the size of label, the position of label, data elements and format, working band of frequency etc. With the improving of the standard system, the series of standards for RFID system have gradually solved the following three questions: First, air interface protocol standards of ISO work in five bands of frequency. The problem about the label's function, operation mode and storage are different when it is working in distinct bands of frequency; Second, the problem between application programs. It has maintained the interactive mode between information of RFID and application programs. Third, the application programs needs to support a variety of modes of information collection, but RFID tags have constraints about storage facilities and power dissipation, it has achieved the balance about the two issues.

### (2) EPC

EPC Internet of Things architecture proposed by the EPC-global, it's main part is the Electronic Product Code (EPC), which is a kind of digital product code normal and global, the standard also have developed the air interface protocol and read-write Control protocol, can provide a unique identifier for all the items. EPC works in UHF band which is from 860MHz to 930MHz, and uses 96-bit encoding. In particular, EPC has set relevant standard about sharing information which includes: EPC middleware specification, Object Naming Service (ONS), the physical markup language (PML). The characteristics of this agreement is that users can flexible exchange and share information (the information includes coding information of sensor and historical information.) by Electronic Product Code Information System (EPCIS). Of course, the information security issues triggered by sharing of information remain to be resolved. Currently, the strategy of EPC-global is as much as possible to achieve compatibility with series of ISO standards. It is worth mentioning that ISO 1800-6 (the air interface protocol) and structure of EPC coding system and ONS framework can provide complete supply chain standards.

### (3) IEEE 802.15.4(Zigbee)

IEEE802.15.4 protocol is an emerging technology standard with uniform short-range wireless communication technology protocol that can work in the 2.4GHz / 868MHz / 915MHz three different bands. Zigbee protocol is made up with PHY, MAC layer of IEEE 802.15.4 and network, application support layer of Zigbee, its prominent features are low cost of network system, easy implementation, reliable data transmission, a short distance Operation, low power consumption, security of each layers, etc. The

standard regards low power consumption, low rate of data transmission, low-cost as its key target, aim to provide a unified standard for the low-rate wireless connectivity. In addition, Zigbee also has the function of self-organized and self-healing network. The summarization and comparison of three international standards is shown in table1.

**Table 1.** Summarization and comparison of these three international standards

Name of standards Comparative items	ISO/IEC18000	EPC	Zigbee
Base on the technology of	RFID	RFID	WSNS
Field	Including basic standards and application standards, involving the entire RFID system	Focus on electronic product code standards	Focus on communication protocol of sensor network
Characteristic	Taking into account the common requirements of applications	Equal to application standards for specific field	A simple application standard for sensors and primary control-oriented
Advantage	The standard not only to ensure the RFID technology has intercommunity and interoperability but also give consideration to the characteristics of application, can satisfy the specific requirements of different applications.	Have solved the transparency and traceability issues of supply chain, provide data sharing.	Protocol is simple and internationally, interactivity and maintainability, super-frame in standard make the protocol has low power consumption, use security mechanisms to ensure security.
Issues to be addressed	Only consider the Auto-ID and data collection standards, do not provide for how to process and share data after data collection.	Security issues brought by information sharing have to be resolved, and the integrality of the entire system have to be perfected gradually.	The application of this standard is not universal in IoT, suitable for use in the field of automatic control and remote control.

### 3 Suggestion on the Establishment of Standards for IoT in China

#### 3.1 A Few Key to Be Considered during Researching on the Standards of IoT

At present, China has not officially released the relevant standards of IoT, but the work has entered to the preparation stage, according to the study of the national standards above (Table 1), our development of the relevant standards of IoT should base on ISO / IEC, EPC-global, IEEE802.15.4. We should reference to the various regional standards which have been put in practice, combined with industry's actual situation of IoT in our

country to develop independent and innovative standards for our country. Also, it needs to consider about both the industrial application development and the compatibility with international standards. Therefore, the following points should be considered.

- 1) The choice of operating band; RFID system of the Perceptive layer of IoT is mainly to solve the problem about the frequency working in the distance from 1 meter to 100 meters, it is recommended to use the bands of 860-960MHz and 2.5GHz. 2.45GHz is ISM band which is an open band, this band exists more than one system (wireless LAN, broadband metropolitan area network) who are influencing mutually, therefore, RFID system are universally working in the frequency of 860-960MHz, but the band of 860-960MHz has been occupied by GSM and CDMA. In general, it must choice the least influenced band in 860-960MHz band. The choice of the specific frequency band should be choice according to large number of tests and applications.
- 2) The research of coding standard; In the process of standard-setting, the first step should test and verify the coding rules which is set by ISO/IEC or EPC, it should focus on improving the code's capacity of error correction in the next step, based on above research to develop a more reasonable and reliable method of encoding and method of calibration, finally, the coordination and unity to the international coding standard should be taken into account when the coding standards with independent intellectual property is proposed.
- 3) The establishment of data transmission protocol; We should test and verify the international air interface protocols which are related to IoT technology according to the provisions of the working band of RFID in China. The interface protocols between readers and related systems, between the various sectors, need to be standardized, and to be regulated, validated when the standards are established. The data transfer can be determined on the basis of the operating characteristics of IoT in our country's various industries.
- 4) The realization of security and reliability; The security and reliability of information of IoT must be the priority when we establish the relevant standards with independent intellectual property rights. The security issues in IoT are mainly reflected in information collection and transmission and information security in perceptive layer, the security of information transmission in core network and professional work of IoT, these requirements above make standard-setter to raise the treatments such as encrypt, segregate and anti-collision for information on the basis of developed standards.

### 3.2 Suggestion on the Content of Standards

Through analysis of the standards that had been proposed, industries condition of our country and the compatibility with other standards, the standards' content should include the followings:

- 1) Coding standards; Unified coding standards can benefit the scaled development of internet of things in China.
- 2) Air interface protocol; Research the major protocols already existing aboard (such as ISO/IEC 1800 series standards), track the development of related

technology and the trend of national standards, analyze domestic achievements and patents of air interface protocols, establish standards with intellectual property rights on the base of the related industries in our country.

- 3) Format of data transmission; Concern about the international relevant standards' developing dynamic, formulate Chinese RFID format of data transmission, and realize the exchange and integration of information in various application platforms.
- 4) Standards for public services; Start with the entire system, develop the standards on the base of applicative features of IoT and Next Generation Network.
- 5) Middleware standards; At present, the standards of middleware are still in the discussion stage in international, so we should take action initiative, the establishment of standards and the development of middleware products with independent intellectual property rights develop the middleware standards.
- 6) Standard for security of standard; The security issue running through the processing, integration, and feedback of information, China should establish its own information security standards to ensure that national interests are not subject to abuse;
- 7) Standards for related products; Establish the basic standards to ensure the quality and performance of products.
- 8) Testing standard;

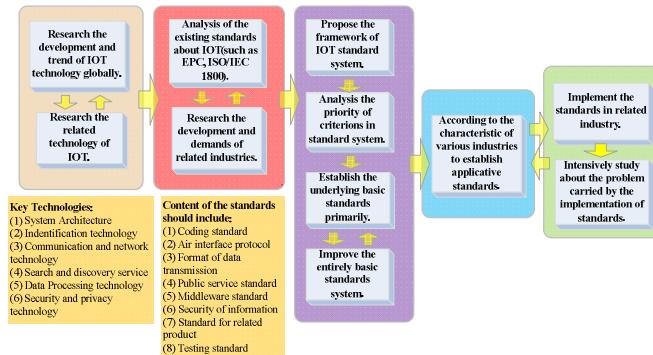
Existed international standards for RFID testing does not involve intellectual property issues, therefore which can be used after appropriate additional modifications base on the characteristics of our applicative bands. Standards for tags, readers, middleware and other products should be developed according to the general specifications of products, at the same time, it need to establish standards for coding, public service system, data protocols and applied technology.

### **3.3 Suggestion on Technology Route of Standardization of IoT**

China has not yet formed unified technical standards for IoT, the factor has became our biggest obstacle to the development of IoT. Therefore, the urgent issues in our country's IoT industry are the unification of standards of IoT. According to the trend of international standards' development, our country's standards for IoT should start with the perceptive layer, According to different application requirements of various of industries and top-level planning, gradually improve the standard system of IoT. And then form a standards system of IoT with independent intellectual property rights and, compatibility with other national standard system.

The route of technology for standards' establishment as is shown in figure1.

The technology involved in IoT is very broad, so the establishment of standards for IoT should have four steps. First, grasp the related technology of IoT globally. Second, fully take into account relevant international standards and China's industrial development of IoT to prove framework of basic standards. Third, develop applicative standards based on the application of the core technology in very field. Last, improve and integrate systematically step by step.



**Fig. 1.** The route of technology for standards' establishment

## 4 Conclusions

Construction of a standardized system is the most important in the development of IoT, China should urgently develop a standard system with independent intellectual property and technology patents which conforms to our development of IoT industry, it is the only way to form the technological core competitive advantage and leading position in IoT industry internationally. Based on the research and analysis in various perspectives such as practical application of the standards which are ISO/IEC1800 series of standards, EPC standards, and IEEE802.15.4 protocols, according to the key technologies and situation as a whole of IoT, this paper analyzes the key points in the process of establishment of standards for IoT and proposes suggested content for relevant standards, thereby, throw out a suggestion on the technique routing of IoT standards, provide theoretical foundation for the development of standardization of IoT.

## References

- [1] Fabian, B., Gunther, O.: Security Challenges of the EPC-global Network. *Communication of the ACM*, ACM 0001-0782/09/0700 (2009)
- [2] Gustavo, R.G., Mario, M.O., Carlos, D.K.: Early infrastructure of an Internet of Things in Spaces for Learning. In: Eighth IEEE International Conference on Advanced Learning Technologies (2008)
- [3] Silverajan, B., Harju, J.: Developing Network Software and Communications Protocols Towards the Internet of Things
- [4] European Research Projects on the Internet of Things (CERP-IOT) Strategic Research Agenda (SRA). *Internetof things—strategic research roadmap (EB/OL)* (September 15, 2009), [http://ec.europa.eu/information\\_society/policy/rfid/documents/in\\_cerp.pdf](http://ec.europa.eu/information_society/policy/rfid/documents/in_cerp.pdf)
- [5] International Telecommunication Union, Internet Reports 2005: The Internet of things. ITU, Geneva (2005)

# The Computational Intelligence of Computer Go

Bin Wu

School of Software Engineering, Beijing University of Posts and Telecommunications,  
Beijing, China, 100876  
wubin19896817@163.com

**Abstract.** Computer Go programs have received some recent attention in AI (Artificial Intelligence) research. One reason is that the Computer Go provides a platform that is both interesting enough to attract people; and complex enough to design the challengeable game AI for successful gameplay. This paper describes AMAF (All-Moves-As-First) algorithm used in the Computer Go. Moreover, we also introduce the improving simulation with domain knowledge. In the end, the experimental result from the Computer Go are presented to demonstrate the effectiveness and efficiency; and the limitation of these methods are discussed, together with possible directions for extending the work towards producing better Computer Go AI.

**Keywords:** Computer Go, AI, AMAF, domain knowledge of Computer Go.

## 1 Introduction

Computer Go has long been considered a difficult challenge in the field of AI and is considerably more difficult to solve than chess [1]. The history of the Game Go could stretch back some 4000 years, and computer Go programs have studied for about forty years [2]. The first computer Go problem was designed by Albert Zobrist in 1968. Although computer Go seems not fresh when compared with modern computer games, it still deserves care analysis. Presently, the best Computer-Go players are at the level of weak amateurs; Go is now considered one of the most difficult challenges for AI, replacing Chess in this role [3].

In the previous study, Monte Carlo methods have been proposed in order to improve the AI of Computer Go. The UCT (Upper Confidence Bound for Trees) and MCTS (Monte Carlo Tree Search) are used in computer Go and the result is considerably successful. However, the evaluation procedure has a little precision; playing the move with highest score in each position does not end up in winning the game. In addition, too many searches for the Monte Carlo Tree are worthless for the final result.

In this paper, we present the AMAF (All-Moves-As-First) heuristic used in the computer Go. It is firstly described by Brügman and further developed by Gelly and Silver. In the computer Go, we observe that some simulations which have different sequences but all moves are the same also have the same result. So the result of one simulation is also significant for other simulation with the same moves. Based on this observation, AMAF updates not only the counts at nodes which update not only the

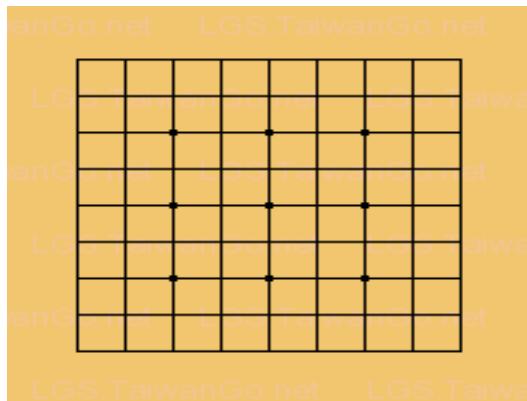
counts at nodes through which the play-out passes, but also many siblings of those nodes [3].

Moreover, we also observe that the pure random simulations always gave meaningless games most of time and sometime the moves which seem to have highest score also lead to fail. Previous study shows that it is more important to get better sequences rather than better moves to make the Monte-Carlo evaluation more accurate [2]. Therefore, before pure random simulation we always verify whether one part of Go broad matches the model which we save before.

The outline of the paper is as follows. In the next section, we briefly describe the computer Go programs. In the Section 3, we introduce AMAF for the computer Go. Section 4 presents the improving simulation with patterns; and the results of some experiments are presented in Section 5. In the end, we talk about the conclusion of this paper.

## 2 The Game of GO

The ancient Asian game of Go, called baduk in Korea and weiqi in China, is played between two players Black and White, who alternatingly place a stone of their own color on an empty intersection on a Go board, with Black playing first. In this paper, we only introduce a computer Go with smaller size board in respect that the level of game search tree is more shallow than the standard board whose board size is 19. Fig.1 shows one kind of computer game GO.

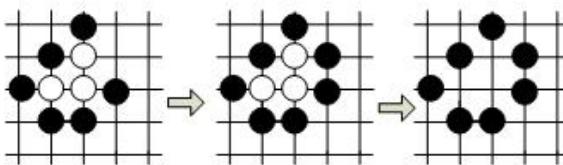
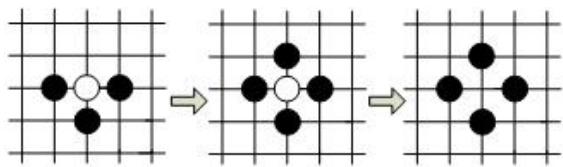


**Fig. 1.** Computer Go

In this paper, we only introduce a set of simple rules which appear in "The Rules and Elements of Go" by James Davies. The detail of the rules can refer to the following.

- The board is empty at the beginning of the computer Go game.
- Player Black moves firstly, then Player White moves. Moreover, a Go move must be played on the intersections of the line.

- At each turn, a player (Black or White) can play his stone of his own color on an empty intersection on the board; or pass his turn at any time.
- A stone or solidly connected group of stones are removed from the board when they are completely surrounded by the opponent, such that no horizontally or vertically adjacent empty point or liberty remains. The Fig.2 shows that the white stones are removed from the board when the black stones surround them.
- No stone may be played so as to recreate a former board position [4].
- The territory of player contains the all the points which his stones are placed on or the territory his stones surround.
- The computer Go ends when one side of players gives up the game or both players pass or both players agree that the game is over.
- At the end of the game, the score of each player is calculated and the player with high score wins the game. The detail of how to calculate the score can refer to [4].



**Fig. 2.** A simple example of the capturing rule

### 3 All Moves As First (AMAF)

The AMAF heuristic is first described by Brügmann in Monte-Carlo computer Go and Gelly and Silver describe the first effort to combine this heuristic with UCT that we are aware of [3]. Generally, the AMAF heuristic is based on assumption that if a move is in a winning simulation, although the sequence is different from the winning simulation; then it will be more likely to be considered a possible move.

In this paper, we introduce the basic AMAF algorithm, which combines UCT with the AMAF update after each play-out. This AMAF algorithm is so similar to the UCT (Upper Confidence Bound for Trees) algorithm; but the only difference between them is in the way of selecting a branch from the trunk and the updating the information of nodes. Detail about the procedure of AMAF algorithm can refer to the following.

- Step 1: Begin search in the root node.
- Step 2: Determine whether the search node is first visited.

- Step 3: If this node has been visited before, then search its child node which has highest score, the score which is calculated by the following equation and go to Step 2; otherwise, go to Step 4.
- Step 4: Do some simulation until the end of the game and get the final result from the simulation.
- Step 5: Update the Monte-Carlo Search Tree based on the AMAF update, which updates not only the counts at nodes through which the play-out passes, but also many siblings of those nodes [3].
- Step 6: Add the new child node based on the degree of search.
- Step 7: Repeat the above 6 steps till the end of simulation time.

## 4 Improving Simulation with Patterns

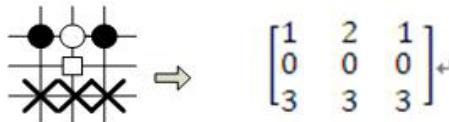
This section talks about Pattern Matching method used in the computer Go. For a computer Go program, the patterns is one of the most important way to save the knowledge about Go. A pattern is a local arrangement of stones on the board which usually represent a local situation. This method is so similar to the human player uses fixed patterns for the game Go.

This section is consisted of 2 subsections. Subsection 4.1 introduces the method that we transform the pattern information. Subsection 4.2 presents the isomorphic patterns in the computer Go game.

### 4.1 Transform the Pattern Information

For a particular computer Go game, our task is to compare each area of the board and find out which of those patterns occur and where they occur. It is hard for our program to do the pattern matching task directly. Therefore, before the pattern matching task, we must deal with the information about the patterns.

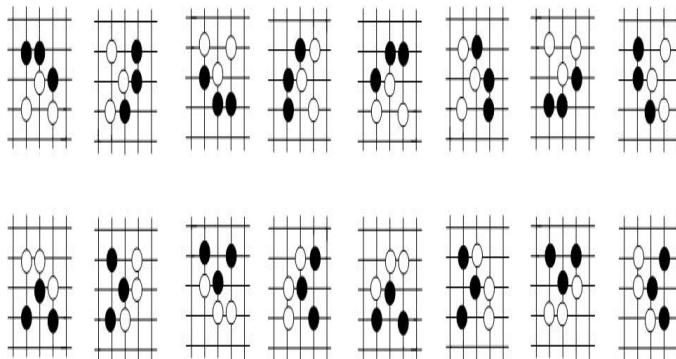
A pattern includes a set of point in this area and the size information. For each point, we use the value 1 to present the black stone, the value 2 to present the white stone, the value 3 to present the point which we don't care about and the value 0 to present the empty position. A simple example can refer to the following.



**Fig. 3.** A simple example of transformation

### 4.2 Isomorphic Patterns in the Computer Go Game

Go is non-directional (isotropic). Whereas pawns in Chess must always move away from their owner, the rules of Go do not assign one side of the board to each player [5]. Therefore, each pattern can be rotated and mirrored; in addition, the color of the stone can be reversed. Therefore, each pattern can be described as 16 forms.



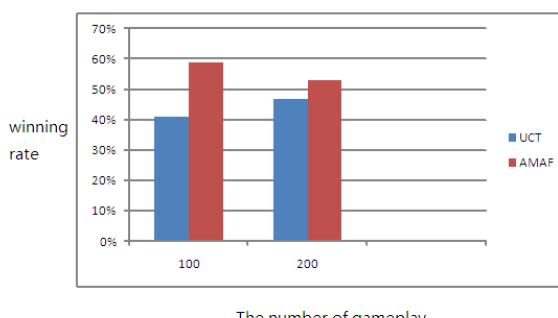
**Fig. 4.** Pattern are invariant to reflection, rotation and color inversion

Figure 4 shows that pattern can be invariant to reflection, rotation and color inversion. Assume that all of the patterns are stored in the memory and a portion of the current board is given, a very simple pattern matching method is available:

- 1) If black is to move, convert the color.
- 2) Compare the array of current area with the pattern array in the memory.
- 3) If a match isn't found, apply the methods to transform the chuck to attempt to match each pattern in the database.
- 4) If one kind of isomorphic chuck match the pattern in the database, the function will return true; otherwise, the function will return false.

## 5 Experiment

In this section, we are going to combine the AMAF algorithm and domain knowledge with the old computer game Go, which is just an UCT Go program. In order to better test these methods used in computer Go game, we conduce an experiment against the old computer Go. We compare the performance of AMAF algorithm by playing Game 100 times and 200 times against the old computer game Go with UCT. The result can refer to the following [6].



**Fig. 5.** Comparison of the performance between the new program with AMAF and the original program with UCT

## 6 Conclusion

In this paper, we proposed AMAF algorithm and pattern matching technique for computer game Go. Our results of simulations have shown that the proposed method works a little well. However, we also observe that, as more simulation time is performed, the performance of AMAF algorithm is close to the UCT algorithm. Moreover, when the number of the patterns is very large, the efficiency of pattern matching algorithm is quite low.

## References

1. Computer Go, Wikipedia, retrieved from,  
[http://en.wikipedia.org/wiki/Computer\\_Go](http://en.wikipedia.org/wiki/Computer_Go)
2. Gelly, S., Wang, Y., Munos, R., Teytaud, O.: Modification of UCT with Patterns in Monte-Carlo Go. Technical Report 6062, INRIA (2006)
3. Helmbold, D.P., Parker-Wood, A.: All-Moves-As-First Heuristics in Monte-Carlo Go. In: IC-AI 2009, pp. 605–610 (2009)
4. Rules of Go, Wikipedia, retrieved from,  
[http://en.wikipedia.org/wiki/Rules\\_of\\_Go#Stones](http://en.wikipedia.org/wiki/Rules_of_Go#Stones)
5. Drake, P., Levenick, J., Veenstra, L., Venghaus, A.: Pattern Matching in the Game of go. The Journal of China Universities of Posts and Telecommunications 14(1), 100–105 (2007)
6. Xie, F., Li, H., Yu, P., Liu, Z.: Progressive Feedback Adjustments in Monte-Carlo GO Simulations. In: Proc. of ISCIS 2009 (2009)
7. Lee, C.-S., Wang, M.-H., Guillaume Chaslot, J., et al.: The Computational Intelligence of MoGo Revealed in Taiwan’s Computer Go Tournaments. IEEE Trans. Comput. Intellig. and AI in Games 1(1), 73–89 (2009)

# Visualizing the Random Forest by 3D Techniques

Min Yang<sup>1,2</sup>, Hexin Xu<sup>2</sup>, Dingju Zhu<sup>1</sup>, and Huijuan Chen<sup>1</sup>

<sup>1</sup> Shenzhen Institutes of Advanced Technology,  
Chinese Academy of Science, Shenzhen, China

<sup>2</sup> School of Software, Sichuan University  
min.yang@siat.ac.cn

**Abstract.** Random forest which contains a set of decision trees is a popular method in data mining. It has the advantages of high accuracy, high learning speed and the ability of dealing with high dimensional data. The decision model from the training process, however, is non-deterministic because of the sampling process. Although we can calculate correlations between different decision trees to infer the performance, it's not comprehensive to non-specialists. So, the goal of this project is to find a way of visualizing the learning process and the final model using 3D techniques. As a consequence, it can help in model selection by visualizing the patterns of different trees in terms of density, similarity and so on. Moreover, it can help users to understand how rules are learnt and then applied in decision making. Finally, it can provide an interactive interface for manual modifications (e.g. pruning).

**Keywords:** Random Forest, Decision Tree, 3D Visualization, Classification.

## 1 Introduction

With the rapid development of computer technology and the internet, computer information retrieval has gradually become the main channel for people to acquire new knowledge. Therefore, the data mining technology has become one of the most significant fields in computer science.

Random forest, as one of the most widely used machine learning algorithms, has become a very common data mining algorithm. Random forest is a combination classifier which is made up of decision trees and letting the trees vote for the most popular class. An early example is bagging, where to grow each tree a random selection (without replacement) is made from the examples in the training set. Another example is random split selection (Dietterich, 1998) where at each node the splits selected at random from among the K best splits. Breiman (1999) generates new training sets by randomizing the outputs in the original training set. Another approach is to select the training set from a random set of weights on the examples in the training set. Ho [4] (1998) has written a number of papers on “the random subspace” method which does a random selection of a subset of features to use to grow each tree. The existing works gradually set up theoretical basis for random forest, and makes its learning ability more and more strong. Because of the advantages of random forest, it has been widely used in data mining, biological information

identification, financial data analysis, enterprise credit analysis and geographic information detection, etc.

Owing to the complexity of the model of random forest, the 3D techniques in visualizing the random forest became an crucial research subject. The 3d visualization technology for random forest can not only enhance people's understanding of algorithm principle and implementation procedure of random forest, but also make the result of the classification more vividly and pellucid. consequently, the promotion of 3D visualization for random forest has great significance in the management, medicine, biological information subject areas and so on.

This thesis makes a 3D-visualization research for Random Forests. To start with, using the training set T, which is the result of repeat random sampling method (bagging) and training set A, the result of random feature selection method to build a single decision tree. Through N times training, we can get N decision trees. Store the data of these trees into file to prepare to use. And then, read the data of trees from files, which means the logic relations among tree's points, and store them into a data structure. Lastly, make the Random Forests into 3D-visualization by OpenGL and QT and get the final picture.

## 2 Related Principle

### 2.1 The Principle of Random Forest

In the year of 2001, Breiman was innovative in obtaining the random forest algorithm which is an integrated classifier. Random forest can do better to tolerate noise and avoid overfitting than single decision tree since it is the combination of bagging algorithm, CART algorithm and the random subspace which make the random forest has a great performance of classification.

Random forest is composed of multiple decision trees. During the process of generating the decision trees, each new training set is drawn, with replacement, from the original training set. Then a tree is grown on the new training set using random feature selection. The trees grown are not pruned. After a large number of trees is generated, they vote for the most popular class. The common element in all of these procedures is that for the kth tree, a random vector  $\Theta_k$  is generated, independent of the past random vectors  $\Theta_1, \dots, \Theta_{k-1}$  but with the same distribution; and a tree is grown using the training set and  $\Theta_k$ , resulting in a classifier  $h(x, \Theta_k)$  where x is an input vector.

For random forests, an upper bound can be derived for the generalization error in term of two parameters that are measures of how accurate the individual classifiers are and of the dependence between them.

- (1) The smaller the correlation between any tree, the smaller the overall generation error rate of the random forest.
- (2) The larger the strength of any individual decision tree, the smaller the overall generation error rate of the random forest. That is to say, increasing the classification strength of a single tree can improve the performance of random forest.

## 2.2 Mathematical Model of Random Forest

Random forest has great performance on classification since it possess powerful mathematical knowledge bases.

Using the margin  $mr(X, Y)$  function to measure the confidence of the random forest in classification.

$$mr(X, Y) = P_{\Theta}(h(X, \Theta) = Y) - \max_{Y \neq j} P_{\Theta}(h(X, \Theta) = j) = j \quad (2.1)$$

Where the where the subscripts  $\Theta$  indicates that the probability is over the  $\Theta$  space. The margin measures the extent to which the average number of votes at X, Y for the right class exceeds the average vote for any other class. The larger the margin, the more confidence in the classification.

The generalization error is given by:

$$PE^* = p_{X,Y}(mg(X, Y) < 0) \quad (2.2)$$

Based on the law of Chebyshev, when the number of decision trees is large enough, generalization error converges to

$$\lim PE^* = P_{X,Y}(P_{\Theta}h(X, \Theta) = Y) - \max_{Y \neq j} P_{\Theta}(h(X, \Theta) = j) < 0 \quad (2.3)$$

Where k is the number of decision trees. From this formula, we see that it will not arise overfitting with the increasing of the trees.

We define the strength of the set of classifiers  $\{h(X, \Theta)\}$  as:

$$s = E_{X,Y}mr(X, Y) \quad (2.4)$$

Where  $E_{X,Y}(\cdot)$  indicate the expectation of  $mr(X, Y)$ . The larger the value of S, the smaller the generation error of the random forest.

Assuming  $s \geq 0$ , Chebychev's inequality gives:

$$PE^* \leq \frac{\text{var}(mr)}{s^2} \quad (2.5)$$

Where,  $\text{var}(\cdot)$  indicates the variance.

Let's assume  $\Theta$ ,  $\Theta'$  are independent with the same distribution, implying that

$$mr(X, Y)^2 = E_{\Theta, \Theta'} rmg(\Theta, X, Y) rmg(\Theta', X, Y) \quad (2.6)$$

Then we can get that  $\text{var}(mr)$  is

$$\begin{aligned} \text{var}(mr) &= E_{\Theta, \Theta'} (\text{cov}_{X,Y}(rmg(\Theta, X, Y) rmg(\Theta', X, Y))) \\ &= E_{\Theta, \Theta'} (\rho(\theta, \theta') sd(\theta) sd(\theta')) \\ &= \bar{\rho} (E_{\Theta} sd(\Theta))^2 \\ &\leq \bar{\rho} E_{\Theta} \text{var}(\Theta) \end{aligned} \quad (2.7)$$

Where,  $\bar{\rho}$  indicates the mean of the correlation, that is to say, the  $\bar{\rho}$  can be defined as

$$\bar{\rho} = \frac{E_{\Theta,\Theta'}(\rho(\Theta, \Theta')sd(\Theta)sd(\Theta'))}{E_{\Theta,\Theta'}(sd(\Theta)sd(\Theta'))} \quad (2.8)$$

And,

$$\begin{aligned} E_{\Theta} \text{var}(\Theta) &\leq E_{\Theta}(E_{X,Y} \text{rmg}(\Theta, X, Y))^2 - s^2 \\ &\leq 1 - s^2 \end{aligned} \quad (2.9)$$

From the formulas considered above, we can get that an upper bound for the generalization error is given by

$$PE^* \leq \frac{\bar{\rho}(1-s^2)}{s^2} \quad (2.10)$$

It shows that the two ingredients involved in the generalization error for random forests are the strength of the individual classifiers in the forest, and the correlation between them in terms of the raw margin functions.

### 2.3 The Merits of Random Forest

As one of the most widely used machine learning algorithms, random forest has the irreplaceable role in classification and so on. Nowadays, random forest has been widely used in data mining, biological information identification, financial data analysis, enterprise credit analysis and geographic information detection, etc. The reason why random forest has such widely use is that it has the following advantages:

- A. Random forest is so far one of the most accurate learning algorithms.
- B. Random forest is made up of multiple decision trees, so it is suitable for parallel processing and dealing with the actual mass data with high performance.
- C. Random forest can do better to tolerate noise, avoid overfitting and deal with training set with lot of missing data than other classification algorithm.
- D. Random forest can estimate the importance of any feature according to OOB data.

## 3 Realizing the Visualization of Random Forest

### 3.1 The Design of Experiment

The realization of this experiment mainly include two process: (a) Using the data from the training set to generate random forest. (b) Visualizing the random forest by 3D techniques.

#### A Realizing the classification with random forest.

Step 1: Applying bagging algorithm to obtain new training samples set T with replacement, from the original training set S. We can get N new training sets by repeating this rule. The N new training sets can be noted as  $T = \{T_1, T_2, \dots, T_N\}$ .

Step 2: we choose n features from the whole original feature set randomly to generate a training feature set. We note the training feature set as  $A = \{a_1, a_2, \dots, a_n\}$ .

Step 3: Growing the decision trees using T and A. Then we store the structural information of the decision trees in the file so as to visualize the random forest by 3D techniques.

### B Visualizing the random forest

Step 1: reading the information of decision trees from the file and mapping it to the intermediate data structure.

Step 2: According to the logical information of the trees, we can calculate the spatial position of any points of the trees using certain rules.

Step 3: drawing the decision trees with OpenGL、QT.

## 3.2 Experimental Data

We conduct experiments to evaluate the effectiveness and efficiency of the proposed algorithm on the UCI data sets. The data sets are made up of landsat satellite data. The data set consist of 4435 training samples and 2000 test samples. Each sample have 36 attributes (4 spectral bands x 9 pixels) and the attributes are numerical, in the range 0 to 255. There are 6 decision classes: 1,2,3,4,5 and 7. There are no examples with class 6 in this dataset they have all been removed because of doubts about the validity of this class.

## 3.3 Experimental Results

### A The Result of Classification

We use the training data to generate 100 decision trees. Then we use the test data to evaluate the model. We show the classification result in Figure 2. As can be seen from the figure, the third class has the most votes, so we regard the test data is the third class which is grey soil.

```
yangmin@ubuntu: ~/workspace/QtSage
Total item: 100
Class:1, num:7
Class:2, num:4
Class:3, num:40
Class:4, num:30
Class:5, num:6
Class:6, num:0
Class:7, num:13
```

**Fig. 1.** The classification result with random forest

### B The Visualization of Single Decision Tree

The result of growing a single decision tree is showed in figure 2. We can see from the figure that, each branch represents a selection process of the attribute and the route

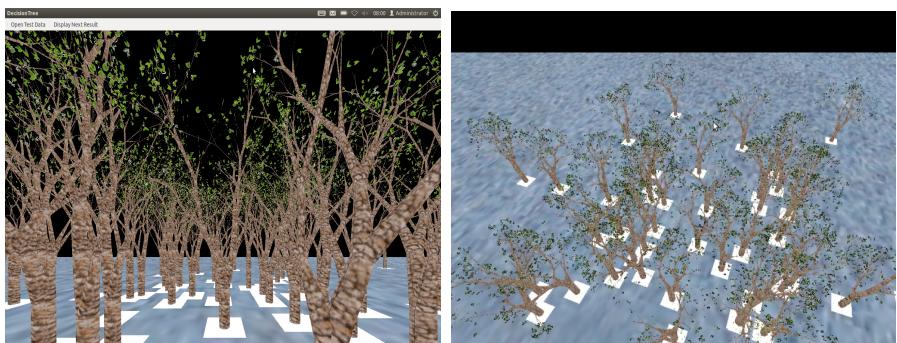
from the root to leaves is a integrated procedure for classification. However, different leaves present different classification results. Namely, if the training data set has  $N$  different kinds of classes, you might see  $N$  different kinds of leaves. Additionally, the distance between the root and the first branch indicates the strength of the corresponding tree which is the criterion to evaluate the decision tree.



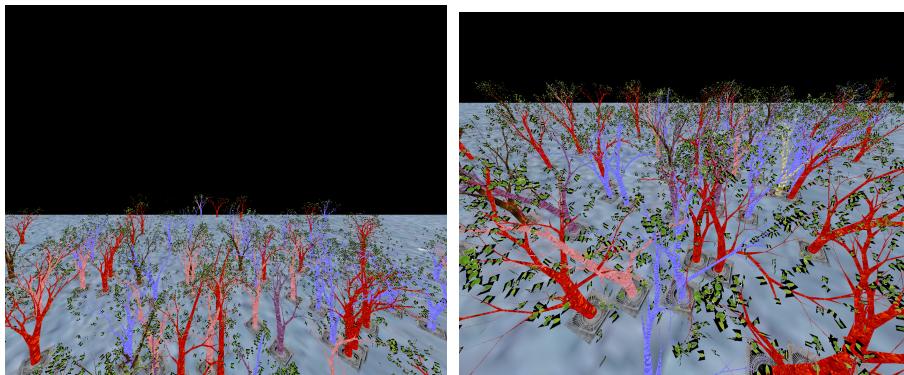
**Fig. 2.** The visualization effect of the single decision tree

### C The Overall Visualization Effect of Random Forest

This system realizes the visualization of random forest. We can see the whole scene of the original random forest in the figure 3 .Using the random forest to classify the test data, we can get the classification and vote results showed in figure 5 and 6. In the forest, the different classification results using different types of decision tree bark represented by various colors and shapes. Using contrast on tree bark, the observer can visually judge in the forest what color and shape of the tree occupies the majority, so as to achieve the final effect of visualization. Additionally, we can evaluate the quality of the random forest model easily. More far apart of any two decision trees, less similarity of the two, and so the total generation error rate of the random forest is smaller.



**Fig. 3.** The visualization effect of original random forest



**Fig. 4.** The visualization effect of after classification

## 4 Conclusion

In this paper, we have succeeded in using 3D techniques to visualize the random forest. The comprehensive presentation provided by the system not only helps people understand the principles behind such an important method, but also provides an easier way to observe the strength of single decision tree and the correlation between the trees which are the criterion to evaluate the performance of the random forest. Detailedly talking, this system has the following major functions: a) showing the random forest with different view settings; b) focusing on one decision tree with a detailed view on tree structures; c) providing a dynamic view on the voting process in classification. Visualizing random forest provides an insight in the interactive model learning and affords the basis for realizing the interactive data mining system.

## References

1. Burges, C.J.C.: A Tutorial on Support Vector Machines for Pattern Recognition. In: Data Mining and Knowledge Discovery, vol. 2(2), pp. 121–167 (1998)
2. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn., pp. 12–19. Wiley Interscience (2000)
3. Communications of the ACM 18(11), 613–620 (1975)
4. Breiman, L.: Bagging predictors. Machine Learning 26(2), 123–140 (1996a)
5. Amit, Y., Geman, D.: Shape quantization and recognition with randomized trees. Neural Computation 9, 1545–1588 (1997)
6. Ho, T.: The random subspace method for constructing decision forests. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(8), 832–844 (1998)
7. Quinlan, J.R.: Induction of decision tree. Machine Learning 1, 81–106 (1986)
8. Kullback, S., Leibler, R.A.: On Information and Sufficiency. Annals of Mathematical Statistics 22(1), 79–86 (1951)

# The Computational Intelligence of the Game Pac-Man

Bin Wu

School of Software Engineering, Beijing University of Posts and Telecommunications,  
Beijing, China, 100876  
wubin19896817@163.com

**Abstract.** The Game Pac-Man is a both challengeable and satisfactory video game that has been the focus of some important AI (Artificial Intelligence) research. The goal for using the Game Pac-Man as a test bed in our experiment is that the Pac-Man Game provides a sufficiently rich and challengeable platform for studying the AI in the computer game, and that it is simple enough to permit understanding of its characteristics. In this paper, we use AMAF (All-Moves-As-First) algorithm to manage the search tree for implementing the Pac-Man AI. Moreover, we also introduce the rule-based policy with the domain knowledge to improve the AI of Ghosts (or opponent). Finally, the experimental result from the Game Pac-Man is presented to demonstrate the effectiveness and efficiency of these algorithms.

**Keywords:** Pac-Man Game, AI, All-Moves-As-First, domain knowledge of Pac-Man Game, rule-based policy.

## 1 Introduction

Pac-Man is a real-time and most popular computer games, which is originally designed by Toru Iwatani for the Namco Company in 1981. In the recent year, as more Artificial Intelligence be concerned, the game Pac-Man has become a novel research topic. This is partly because the Pac-Man Game provides a sufficiently rich and useful platform for developing CI techniques in computer game; and partly because it's simple enough to permit reasonable understanding of its characteristics [1].

The typical version of Pac-Man is a one-player game where the human player maneuvers the Pac-Man character around a maze, attempting to avoid four “ghost” characters while eating dots initially distributed throughout the maze [2]. In this paper, in order to reduce the difficulty of development, the Pac-Man for experiment is a little different from classic computer game Pac-Man. The map is only 16\*16 size and both Pac-Man and Ghosts is only able to make choice of turning direction in the intersection.

In the previous study, the AI of game Pac-Man is implemented by the means of FSM (Finite State Machine). This method is very efficient only when the map of Pac-Man is so small and the rule based on the system is so simple. However, as the map becomes more complex and bigger, it requires the designer to have much hand-code. In addition, there is no existence of meta-programming, planning and looking forward.

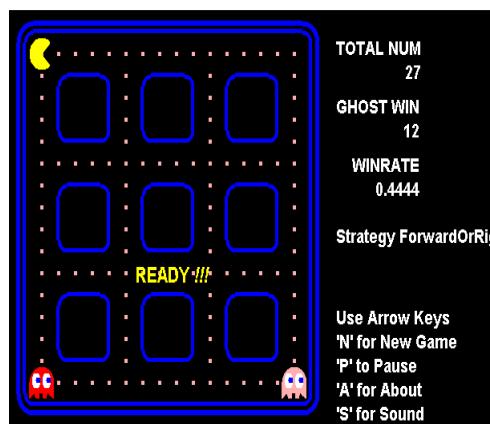
This paper describes some approaches implemented in the Pac-Man game, the approaches which increase the AI of the Ghost. More specifically, this paper presents the AMAF algorithm used to manage the search tree, and discusses about the improving simulation with the rule-based policy. In addition, other aim of our work described below is to explore the feasibility and present practicability of our AI algorithm used in Pac-Man.

The outline of this paper is as following. In the next section, we describe the Game Pac-Man. Sections 3 presents the method we deal with the information of the map. In the Section 4, we will introduce the AMAF algorithm used to manage the search tree. Section 5 talks about the improving simulation with the rule-based policy; and the result of some experiments are presented in the Section 6. Finally, the conclusion of this paper is represented in the last section.

## 2 Description of Game Pac-Mans

Pac-Man (パックマン Pakkuman?) is an arcade game developed by Namco and licensed for distribution in the United States by Midway, first released in Japan on May 22, 1980 [5]. It is the one of the most popular real-time game in the word. In this paper, we use the Game Pac-Man as the test-bed for study the Artificial Intelligence, because it is both challengeable and satisfactory.

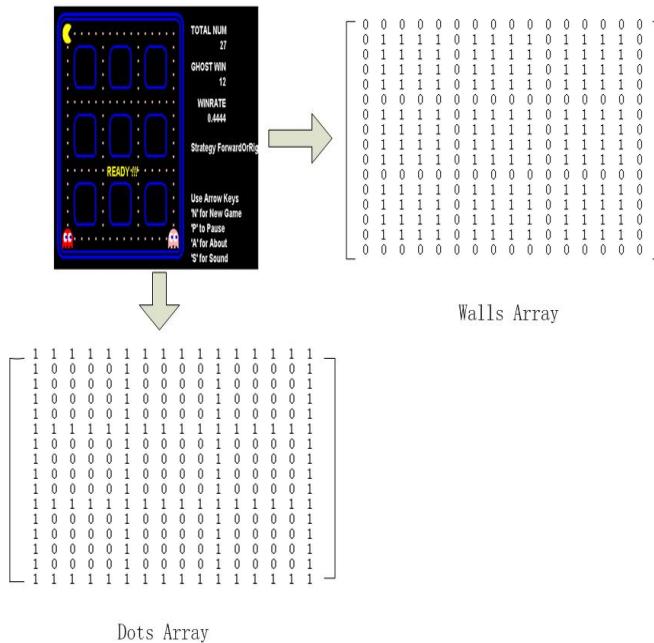
Figure 1 shows our test-bed, Pac-Man game. In the game, the player (or strategy) controls the Pac-Man, which starts in the upper-left of the map. The Opponents (or the Ghosts) which start in left-bottom and right-bottom of the map is controlled by AMAF algorithm with the player strategy information. Our Pac-Man game is the particular Pac-Man game which only contains 112 dots and 2 ghosts. In additioin, both Ghosts and Pac-Man make choice of direction only in the intersections. In any other case, they keep on going in the same direction.



**Fig. 1.** Game of Pac-Man

### 3 Convert the Information of Map

In this paper, we use a JAVA-based implementation for design the game Pac-Man. For a general map of Pac-Man, it is too difficult to present the AMAF algorithm directly. Therefore, we firstly use a two-dimensional array to present the maze information before using them. In this section, we only introduce a simple method to present maze information. This method records the information of every grid in the map.



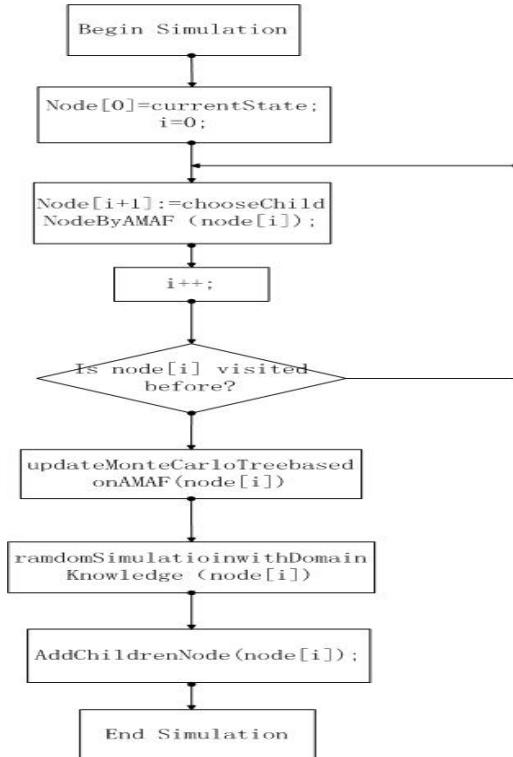
**Fig. 2.** Map information arrays of Pac-Man board

Figure 2 shows that the maze information can be represented by two two-dimensional arrays. In the Walls Array, we use the value 1 to represent the grid with wall; and the value 0 to represent the grids which the Pac-Man and Ghosts can pass through. In the Dots Array, we use the value 1 to represent the grid with dot; and the value 0 to represent the grid without the dot.

## 4 All Moves As First (AMAF)

This section will introduce the AMAF algorithm used in the Pac-Man game. In the previous studies, Artificial Intelligence approaches like MCTS ((Monte Carlo Tree Search), UCT (Upper Confidence Bound for Trees) and AMAF (All-Moves-As-First) are used in the computer Go, but little of research is about these approaches used in the video game just like the Pac-Man game [7]. In this section, we will present the AMAF

algorithm used in the Pac-Man game, which is so similar to the AMAF algorithm used in the computer Go. The only difference between them is the detail of adding children. For the Pac-Man, we only consider the four directions as its children nodes. But for the computer Go, we should add all the moves which are legal as its children nodes. The Figure3 shows the procedures of AMAF.



**Fig. 3.** The procedure of AMAF

## 5 Improving Efficiency with Rule-Based Policy

In this section, we will introduce some rule-based policy for the Pac-Man Game, the rule-based policy which is a set of rules with some mechanism to decide which rule is executed [3]. When compared with the pure simulations with the AMAF algorithm, the advantage of this rule-based policy which can present our domain knowledge is obvious. This is partly because that the rule-based policy can improve the efficiency by saving the simulation time, and partly because that the moves with highest AMAF value sometimes are not the globally better moves. Therefore, we define some rules with highest priority for determining the directions of Ghosts.

This section is considered of 2 subsections. Subsection 4.1 presents the basic definition for the rule-based policy; and in Subsection 4.2, we will construct some rules policy for playing the Pac-Man game.

## 5.1 Basic Definition

Based our domain knowledge for the Game Pac-Man, we define some rules which are easy to complete but has great significance. For a particular situation, we use these rules to match the current state before the AMAF simulation.

In a basic formulation, a rule is a sentence of the form "if [Condition] holds, then do [Action]." [3] Every rule in this paper represents a particular situation in the Pac-Man game, where the AMAF algorithm may be low efficient. In this subscction, we will introduce some basic definition for the rule. These definitions describe the environment information of some particular situation.

- DistanceBetweenGhosts: The nearest distance between the two Ghosts.
- DistanceBetweenGhostandPac-Man: The nearest distance between the Ghost and the Pac-Man.
- KeepDirection: Keep in the direction of last move.
- FromPac-Man: the Ghost go in direction opposite to the Pac-Man.
- FromGhost: the Ghost go in direction opposite to another Ghost.
- StateofCrossroad: Judge whether the Ghost's position is crossroads. If it is a crossroads, then the StateofCrossroads is 1; otherwise it is 0.
- Constant: If all directions are equally good, then Constant is 1; otherwise, Constant is 0.

## 5.2 Constructing Rules for the Pac-Man Game

This subsection describes the rule used in the Pac-Man game. Each rule is described as a condition and an action corresponding to the condition. Its advantage is obvious when compared with the pure AMAF algorithm. In this paper, we only introduce a small part of rules in the Pac-Man, these rules which are simple but can highly improve the AI of the Ghosts.

If DistanceBetweenGhostandPac-Man<2, then FromPac-Man--.

If StateofCrossroad ==0 then KeepDirection++.

If Constant>0 then KeepDirection++

If DistanceBetweenGhosts ==0 then FromGhost++.

The first rule manage ghost capture: if a ghost is too close to the Pac-Man, then the agent should directly catch the Pac-Man. This rule is designed for avoiding unnecessary hesitating. The second rule means that the AMAF algorithm is only used in the crossroads of the map. The third rule is available only when all the directions have the same AMAF value. When all the directions have the same AMAF value, the Ghost will go further in the current direction. The fourth rule is used to prevent two Ghosts in the same cell (when two Ghosts are so close to the Pac-Man), and it is surprisingly effective.

## 6 Experiment

In this section, we first describe the setup of the experiment. The simulated human player plays the Game Strategy A. The detail of Strategy B can refer to [4]. The Opponents or Ghosts are controlled by AMAF algorithm with player's strategy information. In order to better prove the efficiency of our algorithm, the gamers play the game 300 times in the Strategy A, and the result can refer the following.

Simulation Time	Winning Rate
50ms	97.0%
100ms	97.3%
150ms	97.4%
200ms	98.2%
250ms	97.6%

**Fig. 4.** The result of experiment

## 7 Conclusion

In this paper, we proposed AMAF algorithm and rule-based policy for computer game Pac-Man. Our results of simulations have shown that the proposed method works a little well. However, we also observe that, as more simulation time is performed, the performance of AMAF algorithm is close to the UCT algorithm. Moreover, when the game is running on multiplayer online games, these methods are inefficient.

## References

1. Wu, B., Chen, D., He, S., Sun, Q., Li, Z., Zhao, M.: Dynamic Difficulty Adjustment based on an improved algorithm of UCT for the Pac-Man Game. In: Proc. of 2011 International Conference on Electronics, Communications and Control (ICECC 2011), pp. 4255–4259 (2011)
2. Gallagher, M., Ryan, A.: Learning to Play Pac-Man: An Evolutionary, Rule-based Approach. In: Proc. of 2003 Congress on Evolutionary Computation (2003)
3. Szita, I., Lorincz, A.: Learning to Play Using Low-Complexity Rule-Based Policies: Illustrations through Ms. Pac-Man. Journal of Artificial Intelligence Research 30, 659–684 (2007)
4. He, S., Wang, Y., Xie, F., Meng, J., Chen, H., Luo, S., Liu, Z., Zhu, Q.: Game Player Strategy Pattern Recognition and How UCT Algorithms Apply Pre-Knowledge of Player's Strategy to Improve Opponent AI. In: 2008 International Conference on Computational Intelligence for Modelling Control & Automation, pp. 1177–1181 (2008)
5. Pac-Man, Wikipedia, retrieved from, <http://en.wikipedia.org/wiki/Pac-Man>
6. Liu, X., Li, Y., He, S., et al.: To Create Intelligent Adaptive Game Opponent by Using Monte-Carlo for the Game of Pac-Man. In: Proc. of Fifth International Conference on Natural Computation, ICNC 2009, pp. 598–602 (2009)
7. He, S., Gao, Y., Yang, J., et al.: Creating Challengeable and Satisfactory Game Opponent by the Use of CI Approaches. Int. J. Adv. Comp. Techn. 2(1), 41–63 (2010)

# A Reliable AOP Technique Using XML to Handle Semantic Aspect Problems

Eunsun Kim and Byungjeong Lee<sup>\*</sup>

School of Computer Science, University of Seoul  
90 Cheonnong-dong, Dongdaemun-gu, Seoul, Korea  
eskim1208@gmail.com, bjlee@uos.ac.kr

**Abstract.** Utilizing the concept of aspect-oriented programming can modularize the structure of programs clearly. However, the introduction of AOP always does not resolve the code scattering and code tangling of object oriented programming and additional problems may occur. These problems are AOP problems as separation of concerns, pointcut, weaving problem included in study. In this paper we propose an XML based aspect oriented programming technique that it specifies and adjusts the aspects through system intermediary using XML. Our approach can provide reliable dynamic weaving at runtime and indicate the result in case study.

**Keywords:** AOP, XML, Aspect Problems, dynamic weaving.

## 1 Introduction

Developers should enhance readability of program code and minimize change impacts by separating requirements or concerns. The separation of concerns has been achieved through modular programming techniques. Thus, decomposition paradigms such as Object-Oriented Programming (OOP) and Service-Oriented Architecture (SOA) are still dominant in current software engineering.

Despite using OOP, program codes are messy and unreadable. It is also difficult to maintain them because crosscutting concerns such as authentication and logging are tangled with core functional concerns. AOP, introduced later than OOP, is a programming paradigm to handle crosscutting concerns. We can find a way to solve the problems from this paradigm. However, this paradigm is a complementary technique rather than a substitute for OOP.

In AOP developers generate aspects by extracting crosscutting concerns (logging, security, authentication, persistence etc.) from base program. An aspect consists of advice and pointcut, where an advice implements a crosscutting concern and a pointcut defines a location of business code. The aspects are weaved into a joinpoint, where it is a point to locate crosscutting concerns in base program. By applying AOP, developers can build crosscutting concern modules independently without code

---

<sup>\*</sup> Corresponding author.

duplication because they are combined with core concern modules through weaving process.

However, aspect weaving does not sometimes produce the desired results. Many semantic issues related with aspect weaving have been identified [1]. The issues include aspect confliction, vague separation of concerns and the pointcut redundancy. In this paper, we explain some issues which can be caused by introducing AOP and propose an XML based AOP technique to solve the issues.

This paper is organized as follows. In Section 2 we describe AOP issues and related works. In Section 3 we explain the problems caused by introducing AOP and present an XML based reliable technique to solve them. We show a case study about our technique in Section 4. Finally, we draw some conclusions and suggest directions for future work in Section 5.

## 2 Related Works

AOP separates crosscutting concerns into modules or components and combines them through weaving process. However, behavior correctness of AOP application may not be ensured. Especially, when aspects are weaved, there is a possibility that conflicts between aspects occur. For example, it is known that semantic interferences can occur when an aspect for encrypting information in combination with an aspect that finds and removes inappropriate words in text is used [2]. As aspect interference is one of practical problems, some techniques and studies for detecting aspect interactions have been conducted and developed. These studies manage aspect priority and dependency to solve this problem. The ways for resolving aspect confliction and aspect interference have been proposed.

AOP language bindings with implementation framework are generally available, which allow you to directly use AOP with several programming languages such as Ruby, Squeak, Smalltalk, C, C++, Java and so on. However, some web application technologies using scriptlet language are not ready for AOP, such as Java Serve Page (JSP), Personal Home Page (PHP), Active Server Page(ASP), etc [3]. It proposes a more intelligent AOP framework for scriptlet web language that provides (1) two-phase weaving feature; (2) JBoss AOP based cooperation; (3) xml metadata for intelligent aspect mediation. But this situation may invoke the transaction problem in weaving.

A framework for dynamic aspect weaving has been presented as a solution for operating system evolution [4]. The framework has two desirable properties: easy use and fine-tunable function. The intent of this framework is to build an easily extensible dynamic weaving mechanism for operating system evolution. In addition, the system evolution can be achieved without modifying source code. Also, the concept could be realized on many platforms because the desirable mechanism of loadable kernel module (LKM) is widely supported. However, the framework has a limitation that AOP paradigm is not supported in terms of programming aspect.

Many semantic problems, referred to as Aspect Weaving Problems (AWPs) could be introduced in the aspect weaving process [5]. This study proposes a taxonomy of

the AWPs to clearly define and classify the problems. Based on the taxonomy and unified modeling approach to detect the AWPs, a conceptual analysis framework for the AWPs is also proposed. However, the framework is likely to be complex and heavy because of dealing with all the weaving for each language.

A fragile pointcut problem can be tackled by declaring pointcuts in terms of a conceptual model of the base program, rather than defining them directly in terms of how the base program is structured [6]. This study achieves an effective decoupling of the pointcuts from the base program's structure. However, as programmers should learn the language to describe in conceptual level, their burden increases. Because they should also use elements in conceptual level for defining aspects, AOP integration problems are occurred.

An aspect join point model has been proposed, which focuses on the relations between aspects [7]. The model extended existing join point models that focus on the relations between aspects and the base program. This study also extended AspectJ to incorporate this extended join point model. The resulting language, Oarta, can manage interactions between aspects and treat aspect composition as an aspect itself. However, this study does not cover order problem for dependency.

### **3 Aspect Problems and Reliable AOP Technique**

Traditional OOP may have undesirable problems such as code scattering and code tangling in software development. By separating crosscutting concerns by aspect and modularizing program structures, AOP paradigm has been introduced as a new way to solve the problems. However, the introduction of AOP cannot resolve completely the problems and additional problems may be occurred. In this section, we show aspect problems such as separation of aspect, pointcut problem and weaving problem.

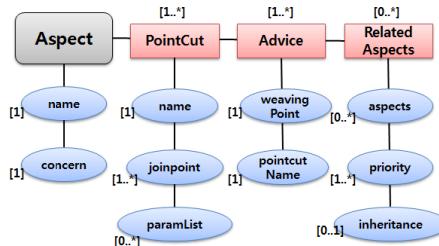
The basic key of AOP is separation of concerns. Separating of functional concerns and nonfunctional concerns clearly makes it possible to reuse and modularize program. Unless separation of functional concerns and nonfunctional concerns are accomplished, code dependencies increase [8].

One aspect should be related with one concern and an advice should be clear. For example, if one aspect is related with different concerns, dependencies between modules increase and program is largely affected by some modification. If this kind of aspect is also simultaneously weaved at the same joinpoint, programmers are likely to fail to produce desired results.

Programmers should define and use pointcuts well. However, pointcut definition problems may be occurred when they define aspects with carelessness. For example, a problem is that a pointcut does not match a joinpoint. This problem is possible because they can specify a wrong method as a joinpoint or confuse with the other method. Another problem is also that pointcuts are redundant, where we define an aspect and a pointcut by duplicating an existing aspect's pointcut. It is related to requirements duplication.

Finally, aspect interference problem occurs when one aspect disables or changes behavior or applicability (i.e, weaving with the base system) of other aspects. And if a

number of aspects are weaved at the same joinpoint simultaneously and the execution order of their advices has not been defined by programmers, the weaving can cause unexpected behavior of program.



**Fig. 1.** Aspect specification schema

Therefore, we propose a reliable aspect oriented technique to solve these aspect problems, where we consider requirements such as decoupling, customization, and reliability. We specify aspects in XML file and weave them into system at runtime (Fig. 1). By locating an intermediary between XML file and application, we support reliable weaving that verifies each attribute of aspects specified by developers.

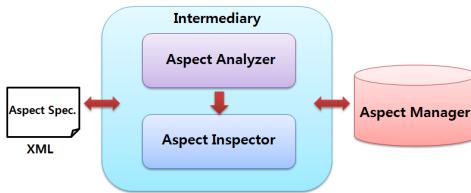
We place the intermediary between aspect specification file and web application for ensuring the execution of web application consistently. If developers specify aspects in a XML file, the intermediary checks and verifies the specification.

The intermediary consists of the ‘Aspect Analyzer’ and the ‘Aspect Inspector’(Fig. 2). The Aspect Analyzer investigates and extracts the aspect specification XML file into elements. Then it sends this information to Aspect Inspector. Aspect Inspector verifies the elements to check if the problems occurred. In particular, we perform an internal process to avoid the problems such as pointcut duplication and aspect interference. If the aspect specification does not meet the requirements during the processing, the intermediary sends a warning message to avoid aspect problems. When the specification passes an internal process, a list of aspects is generated depending on joinpoint in aspect manager.

Aspect Analyzer receives aspect XML data, extracts elements from the specification and sends them to Aspect Inspector. Aspect Inspector verifies whether the joinpoint exists in classes using information transferred. If Aspect Inspector does not match any joinpoint, it allows developers to re-specify the aspect using warning message. If there is no problem, Aspect Inspector checks the duplication of pointcut names. This can be identified based on the list of aspects.

In the pointcut name duplication, Aspect Inspector checks whether an advice is specified at the same joinpoint and weaving point. The priority attribute helps to automatically get the final order of corresponding joinpoint from the aspect list without developer’s intervention. Developers can adjust an advice order by using ‘priority’ property of ‘Related Aspects’.

Aspect Manager supports the aspect weaving and manages aspects based on joinpoints. A joinpoint has various types of advice and helps to output desired result. When aspects are added or removed at runtime, Aspect Inspector handles the changes.

**Fig. 2.** Intermediary and Aspect Manager

## 4 Case Study

Recently web services have broadly been used over internet. In this study, we implemented a prototype of web service application. In this web application, web services defined by WSDL are called when users request atomic services such as weather, news, book and so on. When the application handles the web services, cross-cutting concerns exist throughout a system. Thus we separate cross-cutting concerns and translate them to aspects.

We separated crosscutting concerns such as ‘log’ and ‘exception handling’ aspects and implemented them in XML files. And we used ‘jxl’ to manage the aspect information.

In ‘book’ web service, if users input ‘titanic’ and select ‘title’, then information of ‘titanic’ book is showed in table normally. However, the aspect problems may occur when aspects are applied to the application at runtime. We will show how to solve the aspect problems by using our technique.

First, when an exception occurs, we carry out the aspect weaving dynamically to certify the exception handling using XML(Ex. 1). Fig. 3 shows that users did not enter data in inputparameter field, and invoked ‘book’ service. Aspect ‘SeRuntimeException’ is called because users did not input any data in inputparameter field. This aspect has been inserted at ‘commandExecuteAtomicService’ method that called an atomic service with serviceID. We can see that aspects implementing crosscutting concerns are applied using XML. This aspect has two pointcuts, but ‘ServiceAgentActionExp’ pointcut only was applied on this call. The aspect has no relation with other aspects.

```

<Aspect name="SeRuntimeException" concern="Exception Handling">
<PointCuts num="2">
<PointCut name="ServiceAgentActionExp" joinpoint
="kr.ac.uos.se.action.ServiceAgentAction.commandExecuteAtomicService">
<paramList num="2">
<prm>command</prm>
<prm>param</prm>
</paramList>
</PointCut>
<PointCut name="CompositServiceExp" joinpoint
="kr.ac.uos.se.service.core.CompositService.containsOutputParam">
<paramList num="1">
<prm>param</prm>
</paramList>
</PointCut>
</PointCuts>

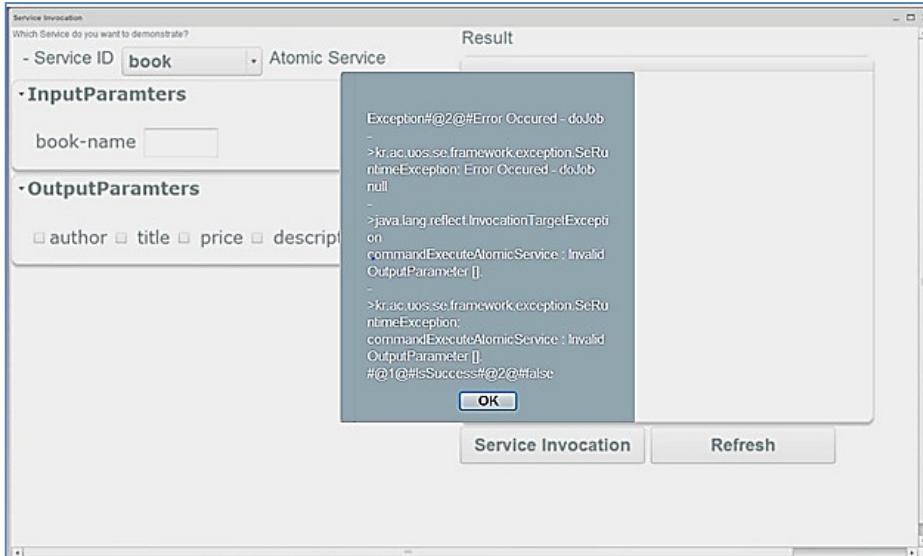
```

```

<Advices num="2">
<Advice weavingpoint="around" pointcutName = "ServiceAgentActionExp"/>
<Advice weavingpoint="around" pointcutName = "CompositeServiceExp"/>
</Advices>
<RelatedAspects aspects="" priority ="0" inheritance="N"/>

</Aspect>

```

**Ex.1** XML code of SeRuntimeException aspect**Fig. 3.** SeRuntimeException aspect result

As mentioned in Section 3, pointcut problems are related to requirement duplication. Therefore we should check duplicated pointcut names.

We defined the ‘MessageLog’ aspect to apply ‘MainActionModel.clientRequest’ joinpoint. Aspect ‘MessageLog’ records the log that services are called and its pointcut name is ‘RequestLog’. But the pointcut name ‘RequestLog’ has already been used in aspect ‘XMLLog’ that records the service called in XML. This is duplicate pointcut name. Aspect confliction can be occurred by duplicate pointcut name. The intermediary checks the information of aspect specification from Aspect Manager. It extracts the aspect’s elements from aspect specification and compares the elements with each other. If a pointcut name is redundant with other pointcut name stored, we sends a warning message(Fig.4).

We also make the log aspect that creates the text file and writes the service invocation. In OOP, we usually create a child class as an extension of a parent class. However, aspect confliction may be occurred when aspects are applied to a method with the same name in a parent class and a child class. We solve this aspect confliction problem between classes through the intermediary. The class ‘ActionModel’ and ‘ServiceAgentAction’ are in an inheritance hierarchy. The ‘ServiceAgentAction’ class

extends the parent class ‘ActionModel’. However, the confliction can occur because of the inheritance relationship between classes when an aspect weaves. Therefore, the intermediary checks ‘inheritance’ attribute of an aspect. If ‘inheritance’ value is ‘Y’, the aspect of child class is called (Fig.5). Otherwise, an aspect of parent class is called (Fig.6).

```

Tomcat v6.0 Server at localhost [Apache Tomcat] C:\Program Files\Java\jre6\bin\javaw.exe (2012. 4. 24. 오후 11:43:20)
정보: JK running ID=0 time=0/22 config=null
2012. 4. 24 오후 11:43:21 org.apache.catalina.startup.Catalina start
정보: Server startup in 953 ms

#####
Request -
Command : commandRetrieveAtomicServiceByKeyword
Parameter : commandRetrieveAtomicServiceByKeyword
    KEYWORD :

#####
Request -
Command : commandRetrieveCompositeServiceByKeyword
Parameter : commandRetrieveCompositeServiceByKeyword
    KEYWORD :
class kr.ac.uos.se.framework.db.DBAccessManager :: DataSource instance..... org.apache.tomcat.dbcp.dbcp.Basic
ActionModel Information...
ActionModel Information...
Logging...
Logging...
#####
Duplicated PointCutName!!!!
ReWrite PointCutName!

```

**Fig. 4.** Duplicated pointcut name message

```

+*****+ServiceAgentAction*****+
Command : null
RequestParameter : null
ResponseParameter : kr.ac.uos.se.framework.collections.ResponseParameter@69a4cb

```

**Fig. 5.** Aspect log result of child class

```

*****+ActionModel*****+
Command : commandRetrieveCompositeServiceByKeyword
RequestParameter : kr.ac.uos.se.framework.collections.RequestParameter@1415056
ResponseParameter : kr.ac.uos.se.framework.collections.ResponseParameter@1014e21]

```

**Fig. 6.** Aspect log result of parent class

## 5 Conclusion

Traditional OOP may have undesirable problems such as code scattering and code tangling in software development. By separating crosscutting concerns by aspect and modularizing program structures, AOP paradigm has been introduced as a new way to solve the problems. However, the introduction of AOP cannot resolve completely the problems and additional problems may be occurred.

AspectJ and SpringAOP are well-known AOP tools, but they are likely to be complex or heavy and do not handle order problem for dependency. Therefore, we proposed a reliable aspect oriented technique that solves the AOP problems by using runtime weaving. This technique generates aspects dynamically because of using XML. We translate the information of aspect to the intermediary. After the intermediary analyzes the information as elements, they are saved in Aspect Manager as a list and weaved into web application. It is possible to realize the reliable weaving because aspects are weaved dynamically at runtime.

**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(No.2011-0026461).

## References

1. Kun, T., Cooper, K., Kang, Z.: A Classification of Aspect Composition Problems. In: International Conference on Secure Software Integration and Reliability Improvement, pp. 101–109 (2009)
2. Aksit, M., Rensink, A., Staijen, T.: A Graph-Transformation-Based Simulation Approach for Analysing Aspect Interference on Shared Join Points. In: ACM International Conference on Aspect-Oriented Software Development, pp. 39–50 (2009)
3. Yuan-Chih, Y., Shing-chern, D., Dwen-Ren, T.: An intelligent Aspect-Oriented Framework for Web Application. In: International Conference on Networked Computing (INC), pp. 1–5 (2010)
4. Jing, C., Hui-Min, S., Chien-Fu, C.: An Aspect-Oriented Framework for Operating System Evolution. In: ACM Symposium on Applied Computing, p. 4 (2010)
5. Kun, T., Cooper, K., Kang, Z.: A framework based approach for unified detection of Aspect Weaving Problems. In: IEEE International Conference on Information Reuse and Integration, pp. 132–140 (2010)
6. Kellens, A., Mens, K., Brichau, J., Gybels, K.: Managing the Evolution of Aspect-Oriented Software with Model-based Pointcuts. In: Hu, Q. (ed.) ECOOP 2006. LNCS, vol. 4067, pp. 501–525. Springer, Heidelberg (2006)
7. Marot, A., Wuyts, R.: Composing aspects with aspects. In: International Conference on Aspect-Oriented Software Development, pp. 157–168 (2010)
8. Netiniant, P.: Separation of concerns for multithreads object-oriented programming. In: International Conference on Electrical Engineering/Electronics Computer Telecommunications and Information Technology, vol. 02, pp. 718–721 (2009)
9. Havinga, W., Nagy, I., Bergmans, L., Aksit, M.: A Graph-Based Approach to Modeling and Detecting Composition Conflicts Related to Introductions. In: International Conference on Aspect-Oriented Software Development, Vancouver, pp. 85–95 (2007)

# **Design of a Smart Meter Recorder with Mass Storage Based on DL/T645-2007 Protocol**

Jia Guo and Dong Liu

School of Information Engineering, Information Engineering University of PLA,  
450001, Zhengzhou, Henan, China  
195001925@qq.com, zdnumber@163.com

**Abstract.** A new software and hardware design which make use of C8051F410 MCU to gather and save the parameter of Smart meter. Research and Analysis the Communication Protocol of DL/T645-2007. On basis of this Protocol .we develop a communication software which is maintained with national standards to exchange data with Smart meter .We also make use of the RS-485 circuit to connect the recorder with the Smart meter . Develop the SPI interfaces in MCU to achieve reading and writing the great capacity SD card and the card has been initialized by FAT16 file system .The active power data , reactive power data and shape data are saved in the SD card together. The experiment showed that the design can gather and save a great number of data exactly and reliably .It can be satisfied with the need of testing and analyzing the power.

**Keywords:** C8051F410, Protocol of DL/T645-2007, SD card, RS-485, FAT16.

## **1 Introduction**

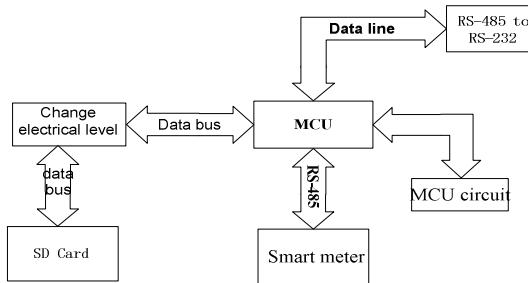
With the development of the society and economy, Constructing intelligent power grid has become a worldwide development trend. AMI(Advanced Metering Infrastructure) as a very important of the smart grid it plays an important role. AMI is composed by the smart meter and data management system in the electric power company also with the communication system. How to get and store the measured data form the smart meter is very useful .It can help us to reduce the line loss manage the user side and operate the power load maximum limit.

The Communication Protocol of DL/T645-2007 which is used to transmit data between the smart meter and the acquisition terminal is the national standards for the smart meters communicating and used widely. With the development of the information technology, SD card as a small volume, low price, high capacity data storage, provides favorable conditions. This paper described how to design the smart meter recorder and introduced the hardware circuit and software. The C8051F410 MCU is used in the equipment as a CPU. The paper also showed how to use C8051F MCU to make embedded development, the design of the communication circuit based

on RS-485, use the high capacity storage SD card, analysis the Communication Protocol of DL/T645-2007.

## 2 Design Scheme of the System

The equipment make use of the RS-485 circuit to connect the C8051F MCU and the smart meter and store the data getting from the smart meter in the SD card, then the data will be removed to the PC and be researched. The equipment also can communicate with the PC. The C8051F410 MCU which produced by Silabs Company in America is used as the CPU. The RS-485 communication circuit which has strong anti-interference ability is used to transfer the data which is gathered by the MCU. The transferred data will be stored by the SD card in FAT16 file system pattern and the gathered data can also be read and processed by the PC. The design of the system is showed as figure 1:



**Fig. 1.** The Design of the system

## 3 The Design of the Hardware

### 3.1 Introduction of C8051F410

The equipment use C8051F410 MCU as CPU .C8051F410 devices are fully integrated, low power, mixed-signal system-on-a-chip MCUs. With on-chip Power-On Reset, VDD monitor, Watchdog Timer, and clock oscillator, the C8051F410 devices are truly standalone system-on-a-chip solutions. The Flash memory can be reprogrammed even in-circuit, providing non-volatile data storage, and also allowing field upgrades of the 8051 firmware . User software has complete control of all peripherals, and may individually shut down any or all peripherals for power savings.

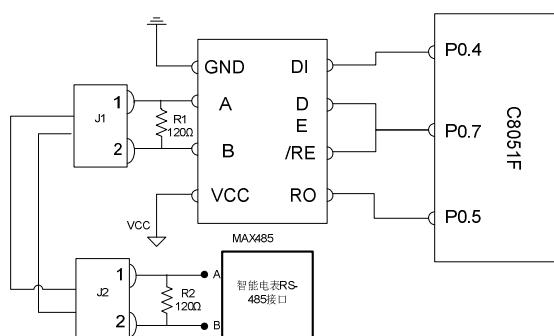
The on-chip Silicon Laboratories 2-Wire (C2) Development Interface allows non-intrusive (uses no on-chip resources), full speed, in-circuit debugging using the production MCU installed in the final application. This debug logic supports inspection

and modification of memory and registers, setting breakpoints, single stepping, run and halt commands. All analog and digital peripherals are fully functional while debugging using C2. The two C2 interface pins can be shared with functions, allowing in-system programming and debugging without occupying package pins.

Each device is specified for 2.0-to-2.75V operation (supply voltage can be up to 5.25V using on-chip regulator) over the industrial temperature range (-45 to +85 °C). The C8051F410 are available in 28-pin QFN (also referred to as MLP or MLF) or 32-pin LQFP packages.

### 3.2 The Design of the Communication Circuit

With the Data-Gather-System using more and more widespread, the communication between MCU to MCU, MCU to PC, is becoming more and more important, so that each part of the system can reach their respective advantages and realize data transmission. RS-485 is the American Electrical Industry Federation (EIA) that make use of the balance for transmission line twisted-pair multi-point communication standard. It uses differential signal transferring the data; The maximum transmission distance can reach to 1.2 km; It can connect 32 receive-send instrument in all; The minimum sensitivity of the receive instrument can achieve to  $\pm 200$  mV; The maximum transmission rate can reach to 2.5 Mb/s. In this method, the transmission distance is long, the transmission speed is high, the lost and cost is low, the reliability is high. This communication stand is compatible with the Communication Protocol of DL/T645-2007. MAX485 is produced by the Maxim Company which is a kind of interface chip. It is used to connect to the RS-485 chip. It works at the source of +5V, and its rated current is 300  $\mu$ A. The working condition is half-duplex communication mode. It can help to exchange TTL electrical level to RS-485 electrical level. In this paper, the design make use of the MAX485 chip to reach the function of RS-485 communication. The circuit schematic diagram is showed as figure 2:

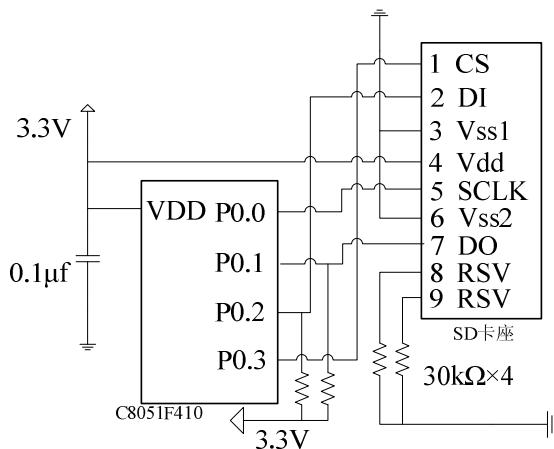


**Fig. 2.** The Circuit schematic diagram of RS-485

### 3.3 The Design of SD Card Memory Circuit

SD memory card is a kind of new large capacity which is non-volatile storage system of external. The equipment need to record and store a lot of gathered data. The SD card has stand pins and can be used wideiy. In this equipment we choose the SD card as the storage devices.

SD card technology is based on MultMedia card . The size of the card is 32mm×24mm×2.1mm. Interface controller is the core of the SD card and the “POWER ON” can provide reset signal. Because of adopting “NAND” technology, power management parts generate programming voltages. The SD card has 3 kinds of working mode, such as:SPI mode, 1 bit SD mode, 4 bit SD mode. There are SPI pins in the C8051F410 MCU, so that we use SPI mode . We can set the “XBRO” and “XBR1” registers that the pins “P0.0, P0.1 , P0.2 , P0.3” can be used in SPI mode . The circuit schematic diagram is showed as figure 3 :



**Fig. 3.** Schematic diagram of the SD card

## 4 Analysis of the Communication Protocol of DL/T645-2007

The Communication Protocol of DL/T645-2007 is used in the equipment to help to communicate between the users and smart meters. It is a very important communication standards. The protocol has changed a lot in fields of phisical layer, data link layer, data identification.

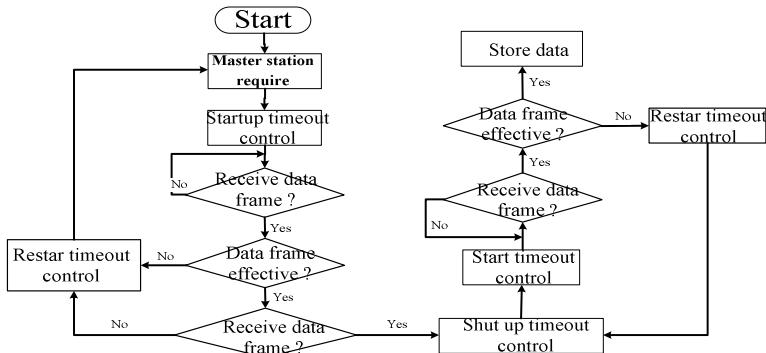
In The Communication Protocol of DL/T645-2007, users can set the address of the meters and each meter has its own address. The number of control command increased to 13. Removing the lowest communication speed 300bps and adding the highest speed 19200bps. If the communication speed has been set, the speed will not be changed. In data logo aspects of data, the data item adopts compressed BCD code.

## 5 The Design of the Software

In the whole software system, it can be divided into gathering data part and storing data part . In this system, the smart meter can communicate with the equipment and store the gathered data in set pattern.

### 5.1 The Design of Communication Software

Because of the Communication Protocol of DL/T645-2007, the equipment use half-duplex communication mode . Each smart meter has its own physical address and use the information frame controlling the communication between the equipment and smart meter. Each frame has a “Start code”, “Address code” , “Data area” , “Control code” , “Length of the data area” , “Check code”. The equipment can send a “Requiring code”, then the smart meter will send some useful information back. The related software flowchart shows as figure 4:

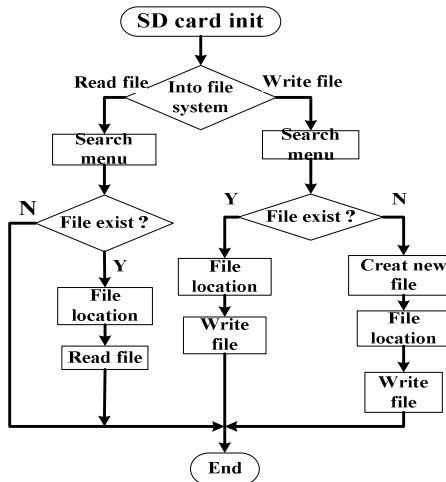


**Fig. 4.** The Flowcharts of RS-485 Communications software

### 5.2 SD Card File System

#### 5.2.1 FAT16 File System

FAT(File Allocation Table) is a kind of table which is used to record the file address. In this system, we use 16-bit space to stand each setor, so we usually call it FAT16. The FAT16 file system contains MBR which is used to record the address of the data and the distance from one memory area to another. SD card can't be divided into several different areas. We can use the MBR AREA recording the size of the based menu, the medium of the discard, the number of the FAT and the capacity of each area. The DIR AREA can be used to record the starting mark and the size of the file. The FAT AREA can record the address of each file. The DATA AREA is used to store the data which gathered from the smart meter. When formatting the FAT16 file area, the size of the cluster is depend on the size of the divided area. Then the number of the DIR, FAT, MBR, DATA will be recorded. The result will be stored in certain area of the DBR memory. The related software flowchart shows as figure 5:



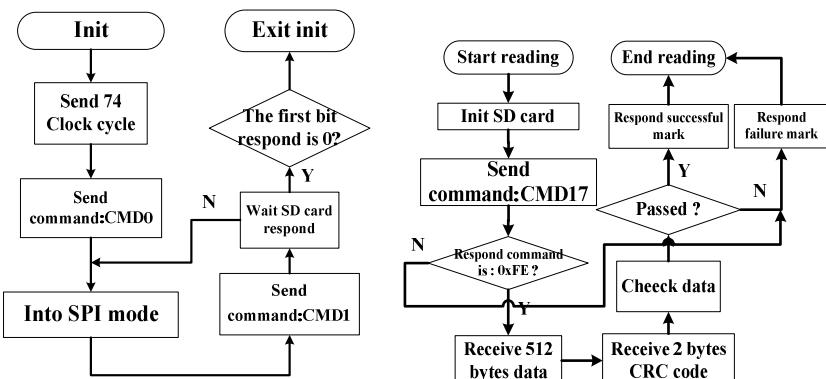
**Fig. 5.** The Flowcharts of the FAT16 file system

### 5.2.2 Initializing SD Card

Following the working rules of the SD card, when the card is connected to the source, the card is in the mode of the bus. At this time, the MCU will send command “CMD0” to the SD card. If the CS is in low level receiving the command, the card will get into SPI mode. If not, it will get into BUS mode. After initializing the SD card, we can read and write the card. All these actions need special command. In the SD card command list we know: The SPI protocol supports operate in single CMD24 part and several CMD25 parts. If you want to read the data which stored in the SD card, you need to get the command CMD17 and the first byte must be 0xFF. After that, the following 512 bytes data can be read. The last 2 bytes is CRC code. Firstly, we need to define the pins on the MCU as follows :

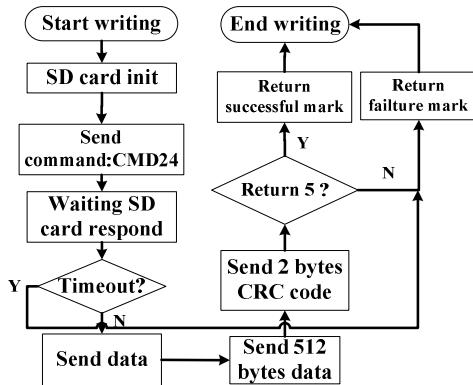
sbit DO=P0^1;sbit DI=P0^2;sbit CS=P0^3;

The related software flowchart shows as figure 6,7,8:



**Fig. 6.** Initialize the SD card Fig

**Fig. 7.** The Flowcharts of reading the SD card Fig

**Fig. 8.** The Flowcharts of writing the SD card

## 6 The End

The experiment of the equipment is successful. It can communicate with the smart meter. The data can be read and written from the SD card successfully. We can also set the working mode if it is needed.

## References

1. Yang, H., Zhang, L.: Study of power load classification based on adaptive fuzzy C means. *Power System Protection and Control* 38(16), 112–115 (2010)
2. Yu, J., Chen, Q.-S.: Research on Means Effectiveness Function  $FP(U,c)$ . *Acta Electronica Sinica* 29(7), 899–908 (2001)
3. Liu, H.-M., Wang, D.-Z., Yuan, X.-L., Chen, X.-Y.: The forecasting of market cleaning price based on TS multi-model. *Power System Protection and Control* 38(23), 33–37 (2010)
4. Communication Protocol of DL/T645-2007. China Power Publish, Beijing (2007)
5. Gerbec, D., Gasperic, S., Smon, I.: Allocation of the load profiles to consumers using probabilistic neural networks. *IEEE Transactions on Power System* 20(2), 548–555 (2005)

# **Research on Temperament Classification Method Based on Fuzzy C-Means**

Wenjun Yang, Yingying Zhu, and Yongxin Li

Institute of Psychology & Behavior, Kaifeng, 450001, China  
195001925@qq.com

**Abstract.** This paper shows that how to use Fuzzy C-Means to classify the temperament and research the Means effectiveness. Raise a method to classify the temperament which only by the figure of the curve and it will not be effected by other complications. Using the actual preferences which provided by the results of the survey to compute the examples .The results shows that the figures which is similar each other are classified together.

**Keywords:** Fuzzy C-Means, Means Effectiveness, temperament, Introduction, Temperament is one of the personal psychological characteristics.

## **1 Introduction**

Temperament is one of the most important psychological characteristics of personalities. It means that, when someone's cognition, emotion, speech, actions, mental activity changes, the power is strong or weak; the changes occurs quickly or slowly and other personality characteristics. Mainly showing the speed of emotional experience, the strength of emotional experience, the implicitness or explicitness of performance as well as the sensitivity or dullness of action, so temperament brings a thick layer of color for all human mental activities.

Understanding the types of temperament has important significance, not only in education but also in training or organization and management. For example, according to the different temperament types of students, educators can teach students in accordance with their aptitude、treat them individually.

In human activity temperament is not decisive, but it can affect the compatibility of some kind of occupation for a person and constitute the foundation of personality at the same time. Temperament does not affect the nature of activity, but can affect the efficiency of activity.

According to the traditional division , temper types can be divided into four types: sanguine, choleric, phlegmatic, depression. However, the amount of people who are classified into typical temperament types is a minority, one person often has multiple temperament types. The paper optimizes temperament type questionnaire, then generates Temperament Survey curve by the questionnaire and get a new temperament type classification by the use of Fuzzy Clustering, which makes the new method of classification is more close to people's true feelings. Not only does the paper analyze

the fuzzy clustering method and study the validity of fuzzy clustering, but also puts forward the temperament characteristic curve based on the method of Fuzzy Clustering. Finally, collected sample questionnaires were simulated classification by the method.

## 2 Fuzzy C-Means

Fuzzy C-Means is a kind of algorithm , which is used to classify the samples. The samples which are classified into one category has the greatest similarity. The samples which are classified into different category has the minimal similarity .

### FCM algorithm using condition[1]

Membership function is a kind of function which represents the degree of object X belongs to set A [2]. We can write it as  $U_A(X)$ ,  $0 \leq U_A(X) \leq 1$ .  $U_A(X)$  means X belongs to set A .

### Basic Idea

Before classifying the samples, the FCM algorithm divide all the feature vector  $X_i$  ( $i = 1, 2, 3, \dots, n$ ) into  $n_c$  classes. Then calculate the clustering center of all the classes. Make sure the membership function and the clustering objective function J which is defined by the distance are minimum. The FCM using fuzzy method classify the samples . The samples will be divided into different categories by using interval (0,1) membership degreeand the similarity degree. According to the fuzzy partition method requirement, the elements in the membership degree matrix U should be in the interval (0,1). Because we should make each classified sample to be normalized . Add all the membership degree together the result should be 1. Just as:

$$\sum_{i=1}^{n_c} u_{ij} = 1, \forall j = 1, 2, \dots, n \quad (1)$$

According to the need , we introduce a Euclidean distance [3]:

$$J(U, X_{c1}, \dots, X_{cn}) = \sum_{i=1}^{n_c} J_i = \sum_{i=1}^{n_c} \sum_{j=1}^n u_{ij}^m d_{ij}^2 \quad (2)$$

In the formula, The U express membership degree matrix ,  $u_{ij} \in (0,1)$  means the j sample for the I category of membership . So we can get the necessary condition of the minimum.

$$X_{ci} = \frac{\sum_{j=1}^n u_{ij}^m X_j}{\sum_{j=1}^n u_{ij}^m} \quad (3)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{n_c} (d_{ij}/d_{kj})^{\frac{2}{m-1}}} \quad (4)$$

### Algorithm Steps

After determining the number which needs to be classified, we use the iterative method to solve the type 3 and the type 4, and thereby get various types of cluster centers  $X_{ci}$  and their membership matrix U[5] are:

- Initialize the membership matrix U and make it satisfy the constraints of type 2.
- Based on the type 3, calculate the cluster center of  $n_c -- X_{ci}, i = 1, 2, 3, \dots, n_c$ .
- Based on the type 4, calculating The objective function value, if it is less than a certain threshold or reach set the number of iterations, stop calculating
- Otherwise, use the type 4 to calculate the new membership degree matrix U, and return to the step 2.

### Algorithm Evaluation

According to the preset value nc, FCM clustering algorithm can get better classification results, at the same time this algorithm is more sensitive for the isolated sample points and special samples can be classified effectively. The paper uses the algorithm in the process of classifying temperament characteristics curve.

## 3 Clustering Validity Function

From the above analysis of fuzzy clustering algorithm, we can see that before classifying the data we must set a classification number in advance, then to classify according to the classification number, however, in fact it is difficult for us to get the number which need to be classified from classification data before classifying. Therefore, we need an algorithm to judge that whether the reset classification number is reasonable or not, namely clustering validity problem.

clustering validity function  $P'(U; c)$  is defined as:

$$P'(U; c) = \frac{\min_{j=1}^c \left( \sum_{i=1}^n u_{ij} \right)}{\max_{j=1}^c \left( \sum_{i=1}^n u_{ij} \right)} \left[ P(U; c) + 1 - \frac{\sum_{j=1}^c \sum_{i=1}^n u_{ij}^2 \|x_i - V_j\|^2}{J_0} \right] \quad (5)$$

Among them, n stands for the number of samples;  $x_i$  stands for sample,  $i = 1, 2, 3, \dots, n$ ; c stands for clustering number;  $V_j$  stands for the jth cluster center,  $j = 1, 2, 3, \dots, c$ ; U stands for the membership matrix.

$$P(U; c) = \frac{1}{c} \sum_{j=1}^c \left( \sum_{i=1}^n u_{ij}^2 \middle/ \sum_{i=1}^n u_{ij} \right) \quad (6)$$

as possible Partition coefficient;

$$J_0 = \sum_{i=1}^n \|x_i - V_0\|^2 \quad (7)$$

as the distance from all samples to  $V_0$  [7]

$$V_0 = \left( \sum_{i=1}^n x_i \right) / n \quad (8)$$

## 4 The Acquirement of Temperament Characteristics Curve

### 4.1 Design Temperament Questionnaire

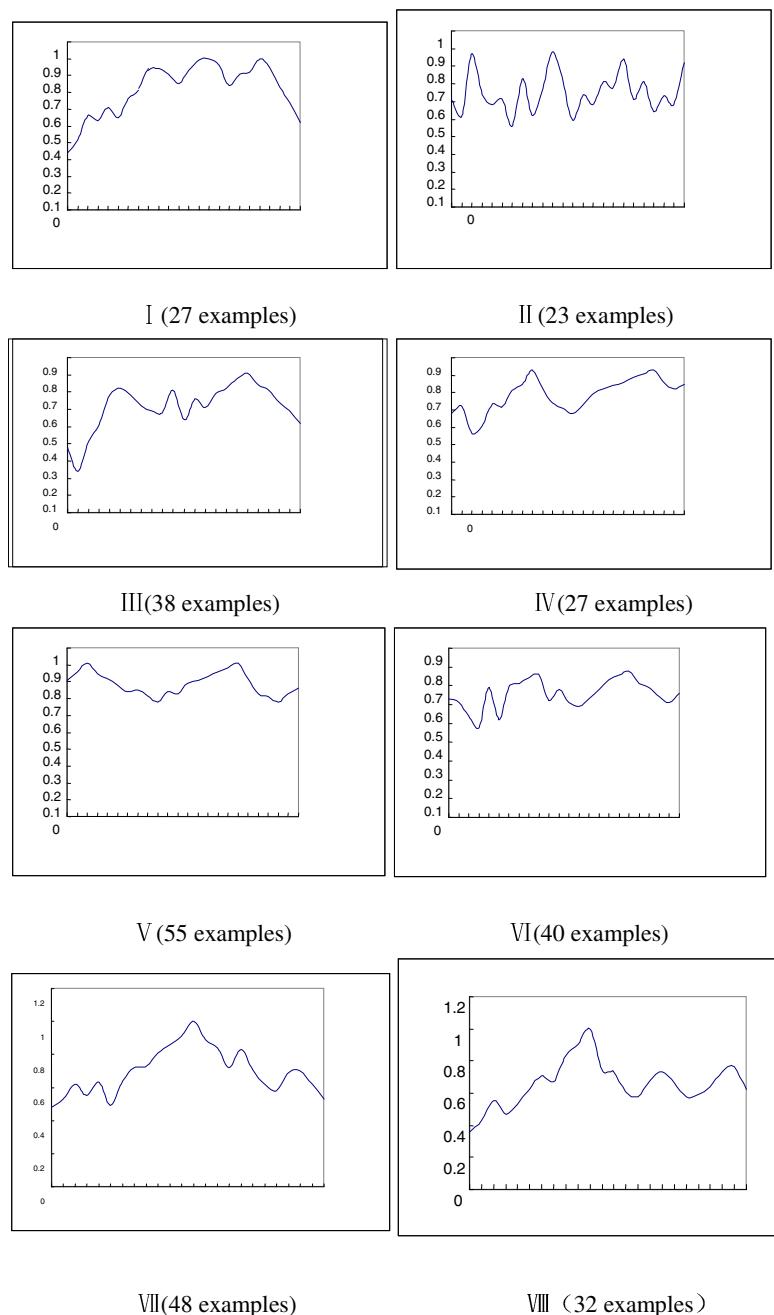
The questions in the questionnaire are still the same as that in STI. According to the recognized degree of respondents, each question is divided into 10 levels from 0 to 1. Then the respondents select one from all options in accordance with their own true conditions, eventually all score points are connected to form the temperament characteristics curve.

### 4.2 The Selection of Respondents

People with the same temperament type will make different choices on the same question, so the respondents in the paper are mainly from students, teachers, civil servants and so on whose working environments are relatively stable.

## 5 Calculation Examples of Temperament Characteristic Curve Classification

In this paper, the datum of this example are from questionnaire results from a middle school of KaiFeng city. These datum are all from field investigation to 300 students of 6 classes, which aren't only true and reliable, but also has a strong representative. The paper classifies the 300 samples' datum by using the fuzzy clustering algorithm. Their temperament characteristic curve and the number of classification are shown respectively in figure 1:



**Fig. 1.** Classified Load Curve

From the classification results of Figure 1 we can see that, compared with our traditional classification methods the classification results of temperament characteristic curve have great differences, by which the temperament types of sample can be divided into 8 categories, each category has similar characteristics. For example, the temperament characteristics of type 1 mainly are grumpy, serious and don't like talking; but not flexible enough, attention is not easy to transfer, follow the beaten track, lack of enthusiasm to career, which belong to mixed type including choleric temperament and lymphatic temperament. Other types all have similar characteristics.

## 6 Summary

In the paper, actual temperament characteristic curves are classified by using fuzzy clustering algorithm and clustering validity function. Compared with traditional classification results, the results we get have great differences. The traditional classification method has been unable to meet the diverse needs of people in modern society.

According to the classification method of temperament characteristic, the temperament types of research objects can have a more intuitive and effective response and its classification is not influenced by other factors. Not only does the paper provide new methods and approaches for us to study the classification of temperament types, but also makes the actual classification results and the feeling of people more appropriate.

## References

1. Yang, H., Zhang, L., He, Q., Niu, Q.: Study of power load classification based on adaptive fuzzy C means. *Power System Protection and Control* 38(16), 112–115 (2010)
2. Yu, J., Chen, Q.-S.: Research on Means Effectiveness Function  $FP(U,c)$ . *Acta Electronica Sinica* 29(7), 899–908 (2001)
3. Liu, H.-M., Wang, D.-Z., Yuan, X.-L., Chen, X.-Y.: The forecasting of market clearing price based on TS multi-model. *Power System Protection and Control* 38(23), 33–37 (2010)
4. Communication Protocol of DL/T645-2007. China Power Publish, Beijing (2007)
5. Gerbec, D., Gasperic, S., Smon, I.: Allocation of the load profiles to consumers using probabilistic neural networks. *IEEE Transactions on Power System* 20(2), 548–555 (2005)

# Author Index

- Ahn, Youngshin 9  
Asogwa, Clement Ogguga 183
- Bai, Hongwei 460  
Banda, Lazarus Obed Livingstone 313  
Bian, Lin 141
- Cai, Ken 597  
Cai, Xuesen 503  
Chang, Chunjie 126  
Chen, Bowen 269, 372, 380  
Chen, Gongliang 242  
Chen, Hao 110  
Chen, Huijuan 639  
Chen, Jiahui 157  
Chen, Jianxin 262  
Chen, Jiawei 439  
Chen, Lin 172  
Chen, LiYun 366  
Chen, Luyu 481  
Chen, Qichao 305  
Chen, Qiyu 409  
Chen, Wenzhi 590  
Chen, Yingmei 254  
Cheng, Xu 394  
Choi, Jaeho 1, 9, 57  
Cui, Baotong 321, 334
- Dai, Xiangyu 64  
Deng, Jian 547  
Deng, Yiqing 453  
Di, Nan 516  
Du, Gang 553  
Du, Wenwen 439  
Du, Yaowei 305
- Fenz, Stefan 565  
Fitarikandro, T. 76  
Fu, Minglei 41, 49
- Gang, Zeng 86  
Gao, Jing 91  
Gao, Yunlu 535, 541  
Geng, Zhenmin 388  
Guo, Jia 660
- Guo, Q.F. 220  
Guo, Xiaohua 352
- Hamed, Ahmed 183  
He, Enze 581  
He, Fei 33  
He, Jie 285, 298  
He, Jinsheng 603  
He, Maolin 41, 49  
He, Ushan 1, 9, 57  
Hu, Chao Ju 486  
Hu, Jinlong 409  
Hu, Jun 216  
Hu, Lei 269, 372, 380  
Hu, Min 466  
Hu, Yan 118  
Hu, Yongtao 535, 541  
Huang, Chun 285  
Huang, Jingyu 285  
Huang, Lei 565  
Huang, Xiaoqing 285, 298  
Huang, Zhen 269, 372, 380
- Inoue, Yoshio 166
- Jeong, Hongkyu 17  
Jia, HongWei 445  
Jia, Shengying 619  
Jiang, Qing 110  
Jiang, Sisi 149, 359  
Jiang, Wen 226  
Jing, Jie 627
- Karim, Awudu 76  
Kharroubi, Fouad 172  
Kim, Eunsun 652  
Kong, Chuncheng 216
- Lai, Dihui 603  
Le, Zichun 41, 49  
Lee, Byungjeong 652  
Li, Chang 466  
Li, Dan 149  
Li, Fei 415  
Li, Guanyu 277

- Li, Guorui 25  
 Li, Hao 627  
 Li, Jiajun 423  
 Li, Jianhua 242  
 Li, Jinchao 529  
 Li, Jingmei 516  
 Li, Lifen 91  
 Li, Lingjuan 262  
 Li, Na 346  
 Li, Tong 254  
 Li, Wen 64  
 Li, Xiaoming 234  
 Li, Xiaoyin 220  
 Li, Xinyu 559  
 Li, Yanchen 611  
 Li, Yang 503  
 Li, Ying 234  
 Li, Yongxin 667  
 Li, Zhen 254  
 Li, Zhihua 388  
 Lian, Yue 141, 473  
 Lian, Yunfeng 366  
 Lin, Risan 305  
 Ling, Xiao 445  
 Liu, Bingwu 572  
 Liu, Chao 394  
 Liu, Chenglong 619  
 Liu, Dandan 192  
 Liu, Dong 660  
 Liu, Dongsheng 611, 619  
 Liu, Li 559  
 Liu, Tao 166, 262  
 Liu, Xinrui 559  
 Liu, Xuan 340  
 Lou, Baodong 200  
 Lou, Xuyang 321, 334  
 Lu, Yu 366  
 Luan, Weiping 104  
 Luo, Bin 220  
 Luo, Jie 298  
 Ma, Nan 423  
 Ma, Qiang 547  
 Ma, Yi 366  
 Ma, Zhiwei 460  
 Meng, Xuejun 70  
 Meng, Zhiqiang 293  
 Nan, Di 509  
 Nan, Min 98  
 Oluyemi, A.M. 76  
 Pan, Wei 220  
 Pei, Changxing 104, 126  
 Peng, SheQiang 401  
 Peng, Yuyang 1, 57  
 Qian, Zhihong 208  
 Ren, Guangwei 481  
 Rong, Xiang 262  
 Shan, Zhilong 118  
 Shao, Guangcheng 200  
 Shen, HongBing 401  
 Shi, Shusu 590  
 Shi, Yimin 277  
 Shibata, Kyoko 166  
 Shiojima, Kouzou 166  
 Si, Mengwei 439  
 Si, Zhengye 149  
 Song, Xin 565  
 Su, Bo 104  
 Sun, Fuquan 394  
 Sun, Lin 529  
 Sun, Mengting 498  
 Sun, QiJin 445  
 Tan, Jiajun 415  
 Tian, Yumin 305  
 Tian, Yun 242  
 Wan, Yunlong 33  
 Wan, Zhou 415  
 Wang, Donghan 572  
 Wang, Honggang 126  
 Wang, Hu 327  
 Wang, Jin 486  
 Wang, Keqiang 597  
 Wang, Shu'en 603  
 Wang, Xiaowen 603  
 Wang, Xing 535  
 Wang, Xue 208  
 Wang, Xueqian 110  
 Wang, Yijun 208  
 Wang, Ying 25  
 Wang, Yongli 226  
 Wang, Zonghui 590  
 Wei, Gan 547  
 Wei, Limin 91  
 Wen, Qiaoyan 581  
 Wu, Bin 445, 633, 646

- Wu, Xianglin 110  
Wu, Yibo 439  
Wu, Ying 529  
Wu, Zushun 149, 359
- Xia, Jing 313  
Xia, Qingsong 346  
Xia, Xiufeng 234  
Xiao, Degui 183  
Xie, Yu 192  
Xing, Changzheng 346  
Xiong, Xin 415  
Xu, Baisen 431  
Xu, Hexin 639  
Xu, Hongyun 157  
Xu, Jingdong 529  
Xu, Liang 293  
Xu, Xiaofei 133  
Xu, Xiaolong 262  
Xu, Yong 98  
Xu, Yuwei 529  
Xu, Zongyu 200  
Xuan, Lijuan 234
- Yan, Lianshan 220  
Yan, Wenzhou 453  
Yang, Fuxing 494  
Yang, Min 639  
Yang, Weidong 431  
Yang, Wenjun 667  
Yang, Xue 388  
Yao, Jingjing 541  
Ye, Qian 321, 334  
Yu, Bowen 494  
Yu, Haihang 553  
Yu, Hongcheng 33  
Yu, Hua 118  
Yu, Jianjun 172  
Yu, Licheng 590
- Zhang, Dawei 248, 394  
Zhang, Dexian 431  
Zhang, Gongxuan 226  
Zhang, Haitao 388  
Zhang, Hengwei 473  
Zhang, Jian 192  
Zhang, Junyong 285, 298  
Zhang, Kehong 523  
Zhang, Ling 409  
Zhang, Liping 611, 619  
Zhang, Qi 516  
Zhang, Shuping 248  
Zhang, Weijun 248  
Zhang, Xianzhong 277  
Zhang, Xiaoming 76, 183  
Zhang, Xuan 327  
Zhang, Xuechi 359  
Zhang, Zhen 192  
Zhang, Zuping 313  
Zhao, Jiantao 141, 473  
Zhao, Pengcheng 98  
Zhao, Yi 9  
Zhao, Zhangji 359
- Zhong, Mei 611  
Zhou, Hong 572  
Zhou, Lin 254  
Zhou, Xiaoping 277  
Zhu, Changhua 104, 126  
Zhu, Dingju 639  
Zhu, Enguo 340  
Zhu, Haiping 33  
Zhu, Lin 157  
Zhu, Ping 133  
Zhu, Shuang 208  
Zhu, Yingying 667  
Zhu, Yongli 91, 460  
Zhu, Yusheng 298  
Zhuang, Yiluan 41, 49