www.zte.com.cn/magazine/English

ISSN 1673-5188
CODEN ZCTOAK

ZTE COMMUNICATIONS

ZTE中兴 ZTE COMMUNICATIONS

An International ICT R&D Journal Sponsored by ZTE Corporation

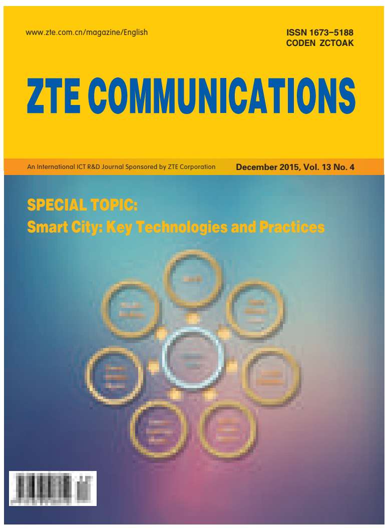
December 2015, Vol. 13 No. 4

SPECIAL TOPIC:
Smart City: Key Technologies and Practices

VOLUME 13 NUMBER 4 DECEMBER 2015

ZTE Communications Editorial Board

CONTENTS



Special Topic: Smart City: Key Technologies and Practices

CONTENTS

ZTE COMMUNICATIONS
Vol. 13 No. 4 (Issue 48)
Quarterly
First English Issue Published in 2003

Supervised by:
Anhui Science and Technology Department

Sponsored by:
Anhui Science and Technology Information Research Institute and ZTE Corporation

Editorial Correspondence:
Add: 12F Kaixuan Building,
329 Jinzhai Road,
Hefei 230061, P. R. China
Tel: +86-551-65533356
Fax: +86-551-65850139
Email: magazine@zte.com.cn

Published and Circulated (Home and Abroad) by:
Editorial Office of
ZTE Communications

Printed by:
Hefei Tiancai Color Printing Company

Publication Date:
December 25, 2015

Publication Licenses:
ISSN 1673-5188
CN 34-1294/ TN

Advertising License:
皖合工商广字0058号

Annual Subscription:
RMB 80

Review

Research Papers

Roundup

Guest Editorial

Jianhua Ma and Weifeng Lv

Smart City: Key Technologies and Practices

Jianhua Ma

Jianhua Ma is a professor in the Faculty of Computer and Information Sciences, Hosei University, Japan. From 1983 to 2002 he researched wireless communications, data encryption, speech processing, multimedia QoS, graphics ASIC, e-learning, CSCW, multi-agents, Internet AV, mobile service, and P2P networking. Since 2003 he has been devoted to what he calls "smart world/hyperworld" with pervasive smart physical u-things or i-things and characterized by ubiquitous intelligence and UbiSafe guarantee. His current research interests include ubiquitous computing, social computing, context-aware service, Internet of Things, wearable technology, digital human clones, and cybermatics. He has published more than 200 papers, authored and/or edited more than 15 books, and has been a guest editor for more than 20 journal special issues. He was the founder and co-chair of the 1st International Conference on Cyber Worlds (CW'02) and was the advisory chair of the 1st IEEE International Conference on Social Computing (SocialCom'09). He has been a founder of the IEEE conferences on Ubiquitous Intelligence and Computing (UIC); Autonomic and Trusted Computing (ATC); Cyber, Physical and Social Computing (CPSCom); Internet of Things (iThings); Smart World Congress (SWC); Cyber Science and Technology Congress (CyberSciTech); and IEEE CIS Task Force on Smart World.

Weifeng Lv

Weifeng Lv is a professor in the School of Computer Science and Engineering, Beihang University, China. He is also the deputy director of the State Key Laboratory of Software Development Environment, China. His research interests and publications span from wireless sensor networks and big data mining and application to large-scale software development methods and smart city. During his career of more than 20 years, he has authored more than 100 academic papers and a university textbook in the field of computer science. He has supervised more than 30 PhD and Master's degree students and won three National Science and Technology Progress Awards. He is now the leader of the "Smart Cities (Phase II)" project jointly supported by the National High Technology Research and Development Program of China and the National Technical Committee on Science and Technology Infrastructure of Standardization Administration of China.

Ubiquitous sensors, devices, networks, and information are paving the way to smart cities in which computation and intelligence are pervasive. This enables reliable, relevant information and services to be accessible to all people. Smart objects, homes, hospitals, manufacturing, and systems will eventually be present in every city.

Although smart city is one of the hottest fields due to its great potential to make our cities more efficient, it is still necessary to clarify the fundamental infrastructures, platforms, and practices needed for truly smart cities. This special issue is dedicated to key technologies and representative practices for building smart cities. Original papers were solicited from smart-city experts, and six papers were selected for inclusion in this special issue. Each paper covers a different aspect of smart city research and practice.

The first paper, "Barcelona Smart City: The Heaven on Earth" by S. Madakam and R. Ramaswamy, covers both a comprehensive review on smart city and a detailed smart city example. Based on extensive data collection and analysis, the authors review smart city origin, concept, research, and applications. The paper describes a representative of smart practice: Barcelona Smart City. The systematic review enables readers to have a clear image about the history and development of smart city. The Barcelona smart city project is also a good reference for other cities in carrying out their smart cities projects.

The second paper, "Smart Cities in Europe and the ALMA Logistics Project" by D. El Baz and J. Bourgeois, first surveys smart city projects in Europe to show the extent of smart transport and logistics, and then describes a smart city project related to a logistic mobile application called ALMA. The application is based on Internet of Things and combines a communication infrastructure and high-performance computing infrastructure in order to deliver high-quality mobile logistic services and that can adapt to dynamic logistics operations.

The third paper, "Smart City: On Urban Operational Collaboration" by R. Cao and W. Kou, expounds the historic origin of urban operational coordination problem that is essential to almost all cities, and then identifies related major challenges and opportunities to make a city smarter. Furthermore, the authors describe the IBM Intelligent Operation Center (IOC) that is a general smart city system framework as an overall solution covering various aspects in implementations of a smart city. Finally, the paper shows a detailed case study using the IOC in building an Emergency Management Centre in Rio de Janerio, Brazil.

The fourth paper, "A Novel Data Schema Integration Framework for the Human-Centric Services in Smart City" by D. Xia, D. Cui, J. Wang and Y. Wang, is focused on the effective scheme to integrate data from various sources and with different characteristics in a city. The authors propose a novel human-centric framework for data schema integration using both schema metadata and instance data for schema matching based on human intervention similarity entropy criteria to balance precision and efficiency. An experiment with real-world dataset has been conducted to

Guest Editorial

Jianhua Ma and Weifeng Lv

test and evaluate the proposed data schema integration.

The fifth paper, "Top-Level Design of Smart City Based on 'Integration of Four Plans'" by J. Cheng and P. Sun, presents a top-level design methodology for smart cities based on the "Integration of Four Plans" covering strategic management, spatial construction, economic development and technical support. The paper also discusses optimal resource allocation; coordination of the development of urban economy, society, resources, environment, and people's livelihoods; and maps out the blueprints for healthy and sustainable development of a smart city. A case study using the proposed methodology for a smart city top-level design is provided.

The sixth paper, "Smart City Development in China: One City One Policy" by B. Wan, R. Ma, W. Zhou and G. Zhang, is focused on the high level policy and development in managing and promoting many smart cities from government's view point. The basic policy is advocated as "One City One Policy" because cities differ greatly. Of ninety cities as first batch of pilot smart cities announced by the Ministry of Housing and Urban-Rural Development (MOHURD), this paper introduces five successful pilot cities (including town and district) as five different models in China's smart city development.

We would like to express our great appreciations to all the authors for their contributions and all the reviewers, in particular, Professor Junde Song, for their efforts in helping to improve the quality of the papers. We are grateful to the editorial office of *ZTE Communications* for their strong support in bringing this special issue to press.

News

Congratulations to the Newly Elected IEEE Fellows

Every year, the IEEE inducts about one-tenth of one percent of the total voting IEEE membership as IEEE Fellows in recognition of their distinguished accomplishments in any of the IEEE fields of interest. This year, four more members of the ZTE Communications Editorial Board became IEEE Fellows for 2016.

Professor Chengzhong Xu, Wayne State University, USA, was named Fellow for leadership in resource management for parallel and distributed systems.

Dr. Fa-Long Luo, Element CXI, Inc., USA, was named Fellow for contributions to adaptive signal processing for hearing and multimedia applications.

Professor Jinhong Yuan, University of New South Wales, Australia, was named Fellow for contributions to multi-antenna wireless communication technologies.

Professor Shigang Chen, University of Florida, USA, was named Fellow for contributions to quality of service provisioning and policy-based security management in computer networks.

The ZTE Communications Editorial Office congratulates them for their achievements and wishes them further success in the future.

DOI: 10.3969/j. issn. 1673-5188. 2015. 04. 001
http://www.cnki.net/kcms/detail/34.1294.TN.20151208.1455.002.html, published online December 8, 2015

Barcelona Smart City: The Heaven on Earth (Internet of Things: Technological God)

Somayya Madakam and Ramaswamy Ramachandran
(National Institute of Industrial Engineering, Mumbai-400087, India)

Abstract

Cities are the most preferable dwelling places, having with better employment opportunities, educational hubs, medical services, recreational facilities, theme parks, and shopping malls etc. Cities are the driving forces for any national economy too. Unfortunately now a days, these cities are producing circa 70% of pollutants, even though they only occupy 2% of surface of the Earth. Public utility services cannot meet the demands of unexpected growth. The filthiness in cities causing decreasing of Quality of Life. In this light our research paper is giving more concentration on necessity of "Smart Cities", which are the basis for civic centric services. This article is throwing light on Smart Cities and its important roles. The beauty of this manuscript is scribbling "Smart Cities" concepts in pictorially. Moreover this explains on "Barcelona Smart City" using Internet of Things Technologies". It is a good example in urban paradigm shift. Bracelona is like the heaven on the earth with by providing Quality of Life to all urban citizens. The GOD is Interenet of Things.

Keywords

smart cities; Barcelona City; Internet of Things; smart mobility; open access data

1 Introduction

Urbanisation is one of the most glaring realities of the 21st century. Cities are growing very fast owing to a large scale urbanisation across the world. Kingsley Davis (1962) explained: "Urbanisation is a process of switching from a spread out pattern of human settlements to one of concentration in urban centre" [1], [2]. In the last two decades, the world has experienced phenomenal levels of urbanization. In the near future, more than half of the world's population will live in cities, and the number of cities with five to ten million inhabitants will continue to rise. About 60 cities will have more than five million people, including cities such as Mumbai, Karachi, Mexico, Lagos, Shanghai, and Beijing [3]. Apart from these, now a days in everybody's mouth, the top most uttering global cities are Vienna, Toronto, Paris, New York, London, Tokyo, Berlin, Copenhagen, Hong Kong and last but not least is Barcelona. What is special about these 10 cities? Recently even the Indian Urban Minister Mr. Venkayya Naidu visited Barcelona to observe the city planning, design, architecture, and urban practises. So what is new in this city? The business giants like CISCO, IBM, Schneider-Electric, HP, Microsoft etc. are always give talk on new urban models by exampling in the case of Barcelona city. Why they always insist chat on particular Barcelona city only? The reason is Barcelona is a Smart City. Let us see Barcelona.

Barcelona was founded by Romans, way back dated on circa 2000 years to its origins as an Iberian village named "Barkeno." With more than 2,000 years of history and a singular identity, Barcelona has always been characterised by its spirit of innovation, enterprise and nonconformity. Cerda, the city council, the Spanish government, civil engineers, architects, and land owners-to strengthen their role in the implementation process and gain control over shaping Barcelona [4]. Thanks to Barcelona to the bold to the adventurous spirit of her architectures, builders and the liberal farsightedness of her planners, is developing into one of the greatest wonder cities of twentieth century [5]. Barcelona, the capital of the autonomous community of Catalonia province in Spain, is now called "Barcelona Smart City." This is the world's first and full pledged converted Brown Field City into Smart or Green Field. In fact this is the Spain's second most populated (around 1.6 million) city. At the same time it is also standing the sixth most populous city in the European Union behind Paris, London, Madrid, Ruhr, and Milan. Barcelona city is well known for its rich cultural heritage. Hands up to the 1992 Summer Olympics, a lot of efforts went into modernizing city while keeping its ancient

This research work is carrying out in partial fulfillment of Somayya's Fellow Programme at National Institute of Industrial Engineering (NITIE), Mumbai. The financial support is fully funding by Ministry of Human Resource Development (MHRD), and the technical guidance from NITIE Professors.

charm. This was considered as one of the best modern Olympic games history. It is predominantly renowned for the architectural works of Antoni Gaudí and Lluís Domènech i Montaner. Barcelona is one of the world's leading economic, commerce, tourist, education, entertainment, media, fashion and Quality of Life centres. Barcelona has a Mediterranean climate. The city has the smallest amount of Green House Gases (GHG) releasing in a newly study documenting how differences in climate, population density and other factors affect GHG emissions in global cities [6]. It is a city of culture, knowledge, business, creativity and wellbeing, pioneering global centre, because it wants to become a role model for Smart Cities.

2 Concepts on Smart Cities

Let us define what does mean by "Smart Cities": Smart Cities are the cities in which they provide Quality of Life to the urban citizens along with economic development, ecologically balance and sustainable for the future generations using complete automated Internet of Things (IoT) technologies. These technologies are generally marry with city sub systems of transportation, security, governance, public utilities like water, waste, gas, power management and other physical infrastructure to bring the operational efficiency. The technologies include Smart Cards, RFID, Quick Response Codes, Electronic Product Code, IPv6, Sensors, Actuators, Wi-Fi, Bluetooth, ZigBee, Near Field Communication, Geographical Information System, GPS, Social Media, Business Intelligence, Ambience Intelligence, Cloud Computing, Tele Medicine, Web 3.0, Big Data Analytics etc.

Giffinger, et.al.(2007), defined that "A Smart City is a city well performing built on the 'smart' combination of endowments and activities of self-decisive, independent, aware citizens." This definition is the milestone in the history of Smart Cities, even though there was academic work and importance given to Smart Cities by Laterasse and Gibson et al.,(1992). Smart Cities are engines for the growth of any national economy. These are cities in which, they think for urban citizens, they do the things on behalf of them; monitor and they control the deviations too. These cities deliver the right solutions for urban dwellers at the right time, at the right place and with right mode 24/7. In the future, these cities are going to talk with people, things and even other cities without any global partiality. These are the cities which functions in 365 days without rest in order to provide Quality of Life (QoL) using Internet of Things (IoT) technologies. In order to provide better life for urban citizens, a huge number of Internet of Things technologies have to be deployed in health, education, transportation, governance, security and utility services. Smart management has to be done in public utilities like electricity, water, gas and waste. Besides technological deployment, other key performance indicators (KPIs) are governance (Central, State, Local), land, environment concerns are required. However, in these cities, first infrastructure will takes place and then people will start to live in, because cities constitute right from scratch. One set of people say that Smart Cities are very Specific, Measurable, Accountable, Relevant, Timely. Some other people say that Smart Cities are abreacted as Sustainable Management Action Resource Tools for Cities. Let we have some explanation about these in **Table 1**.

Table 1. Smart Cities

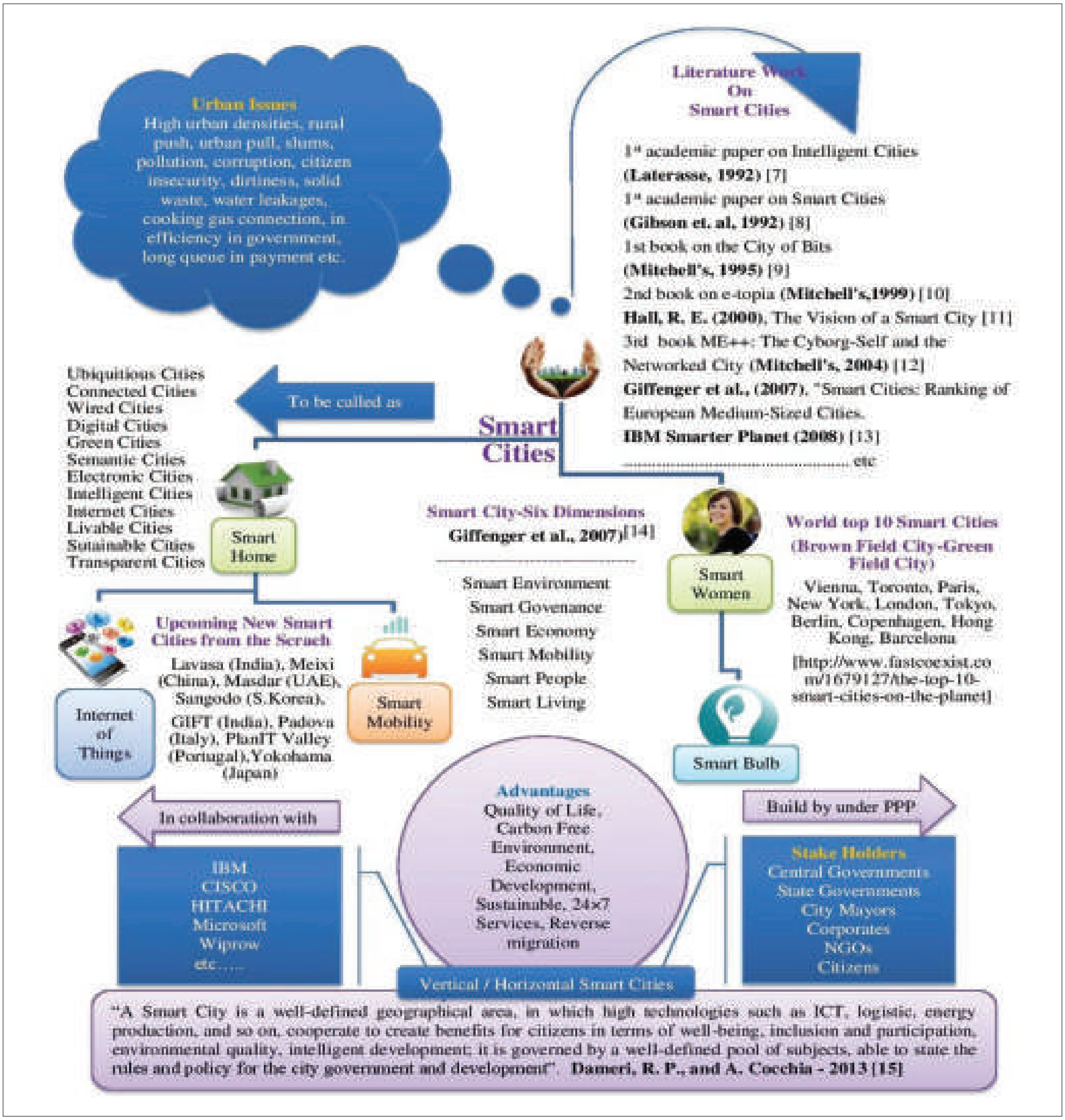
	Explanation 1	Explanation 2
S	**Specific:** All the city functions including mobility, security, governance, public utilities like waste, power, gas should be clearly specify in the software and network. The pre-defined functions, sensors, actuators etc. help to bring right output.	**Sustainable:** These are the cities, generally designed and developed in view of ecologically balanced. Uses renewable energy resources like solar, wind, biofuel, tidal in order to bring carbon free environment and prevent global warming.
M	**Measurable:** These are the cities, in which we can see even consumer power meter at central station. The Smart Grid is one of best bidirectional technology. In this, suppliers also directly find out power leakages, power theft at central level.	**Management:** City with subsystems of smart people, IoT, smart objects (devices) will have unique identities, automation, monitoring and control power. Right things will be done by the right object at the right place at the right time.
A	**Accountable:** The public data will be open. Governance is transparent. Hence corruption by the government will be drastically reduced. These are the cities, in which all the operation of transportation, governance, infrastructure, public utility services should be accountable for the citizens.	**Action:** City Command Control Centers (C4), sitting in the city data centers, will takes action. With the help of smart devices, fully connected City Area Network (CAN), the citizen issues can be solved. There will be solo or inter operable operating systems at each Smart City level.
R	**Relevant:** Because of huge amount of data is generating in zeta bytes by people, devices, objects, with the help of best computing devices and Big Data Analytics, will send relevant data to the concerned object with security mechanisms.	**Resource:** Social: Planners, developers, skilled workers, educators Physical: Roads, buses, railways, rivers, dams, ports Environment: Waste, water, gas, electricity, forest Technological: Hardware, software, networking, IoT, BDA.
T	**Timely:** All citizen services should be in 24×7 in 365 days. On-time transportation facilities for all the city commuters. On-line spot payment systems of telephone, land, tax bills. Automated real time incidents, events, smart health devices. Instantly issuing of date of Birth, Death, other certificates. Real time citizen security using Closed Circuit TV (CCTV) or Internet Protocol (IP) surveillance systems.	**Tools:** Along with hardware, software, networking components and smart devices, the technological God Internet of Things will also plays a vital role in order to get better city operational efficiency. The tools may be sensors, actuators, Wi-Fi, GIS, GPRS, analytics, cloud computing, data centers, web 3.0, Near Field Communication, Radio Frequency Identification, Qucik Response codes, robotics.

3 Literature on Smart Cities

The authors' bird eye view of "Smart City" pheomenon is shown in **Fig. 1**.

4 Methodology

The methodology is critical for any kind of research, analysis, report writing, and publication. In recent years methodology has been increasingly used as a pretentious substitute for method in social, scientific and technical contexts. Research methodology is a systematic way of solving a problem. It is a science of studying how research is to be carried out. Essentially, it is the procedures by which researchers go about their work of describing, explaining and predicting new phenomena.



▲**Figure 1. Author's bird eye view of "Smart City" phenomenon.**

It aims to give the work plan, design and action of research. The beauty of this research is a modern way of data collection using Future Internet (FI) media. The use of Future Internet to aid research practice has become more popular in recent years. In fact, some believe that Internet surveying and electronic data collection may revolutionize many disciplines by allowing for easier data collection, larger samples, and therefore more representative data [16], [17]. The research has been carried out through exploratory study.

4.1 Data Collection

The time taken for data collection is nearly two years. The data collected in a modern methodical way is mainly searched from Google with the key words "city", "smart city", "Barcelona City", and "Barcelona Smart City". The searching was done with the extension format of words (doc, docx, pdf, ppt, pptx). The biggest online database "Google" and "Google Scholar" is the base for lettering this conceptual article Barcelona Smart City. "Knimbus" Indian based online database stands for Knowledge Cloud and is a dedicated knowledge discovery and collaborative space for researchers and scholars. It is also used in searching articles on smart cities. Around 20 YouTube videos are also used for data about some of Smart Cities and Barcelona City. Lots of technical conferences, such as SecureIT-2012 and Smart City: Delivery of Civic Services-2015, workshops, such as IT Innovations for Smart City-2015 and Smart City: India-2015, and symposiums, such as Smart Cities Summit-2014 and TENSYMP 2015: Internet of Things, which are conducted during August 2012-May 2015 to now in India and abroad, have helped us a lot to gather knowledge and share via technical discussions.

4.2 Samples

Since this is an exploratory study, there is no specific sample size. Exploratory research is defined as the initial research into a hypothetical or theoretical idea. Barcelona is a new concept; this phenomenon is trying to prove by several methods of data collection. This data is in different formats including Barcelona city videos, write up articles, city pictures and audios. The authors went for the some samples of corporate top level management video talks and not in traditional in-depth interviews. These are some of interview samples given about Barcelona Cities in different context which talks more in civic services of city:

1) Wim Elfrink, EVP & Chief Globalization Officer (CISCO);
2) Anil Menon, President Smart + Connected Communities (CISCO);
3) Manel Sanroma, CIO, Barcelona City Council;
4) Judith Romera, In City Promotion Director, Barcelona City Council;
5) Toni Vives, Deputy Mayor for Urban Habitat, Barcelona City Council;
6) Mariano Lamarca, Project Leader Smart Cities & Corporative, Wireless Projects, Barcelona City Council.

4.3 Narration

This is a qualitative case study and is an approach to research that facilitates exploration of a phenomenon within its context using a variety of data sources. This ensures that the issue is not explored through one lens but rather a number of lenses that allow for multiple facets of the phenomenon to be revealed and understood [18]. We can go ahead with data analysis using ATLAS.ti or N-Vivo software because data is in qualitative format. However, the analysis of data went thematically after 360 degree level of online data observations. The narration went on some particular writers' and research authors scribbling and speeches. The main themes of description about Barcelona Smart City in this paper includes city open Wi-Fi, Smart Mobility using e-vehicles, Smart Water Management System, Smart Lighting System and last but not least Open Access Data. Apart from these, some dimensions are left without any narration, because of page limit.

5 Barcelona Smart City

The European Commission awarded the European Capital of Innovation ("iCapital") prize to Barcelona (Spain) "for introducing the use of new technologies to bring the city closer to citizens" [19]. In 2008, Barcelona faced challenges as the economy crashed. There were some city mayors, architects, planners, and designers who decided that they wanted to mount city at the global level, which can sustain on par with global cities even in economy crises. They started Internet of Things (IoT) deployment in each and every city dimension for the operational efficiency. The technological advancement also made city planners deploy IoT technologies for better civic services. The Smart City project included Open Data initiatives, offering valuable information to urban citizens and corporate people. The city is providing sustainable growth via resourcefulness on smart lighting, smart mobility and residual energy as well as social innovation. The city is also delivering 'smart services' in a flexible, continuous and smart way through Internet of Things in different parts of Barcelona. As the title of the article "Barcelona Smart City: The Heaven on Earth" implies, the Quality of Life in this city is very high. Here, people are economically rich and with the help of technological GOD (Internet of Things), all the citizen services are providing in round the clock in every corner of the city through City Wi-Fi network. The city is also promoting alliances between research centers, universities, private and public partners through work. Barcelona is a stylish city in Spain with one of the highest densities in Europe. Barcelona is to walk through mile after mile of narrow streets embraced by beautiful old buildings, fronted by small shops. But to hang out in Barcelona is also to taste a form of urban livability almost unknown in North America. People can sit for long hours in some of the best cafes and bars in

Europe, eating some of the best food in the world, and surrounded by a city designed to make the street a second living room [20]. Barcelona is one of the best economic, social and environmentally sustainable cities in the world. **Fig. 2** shows that the Smart City initiatives consisting of Wi-Fi, Open Access Data, Smart Mobility, Smart Water Management System, Smart Lighting System, Smart Waste Management System and Smart Allotment. These technologies are described in detail in the next sections in sequence and demonstrating how they are bringing Quality of Life to citizen.

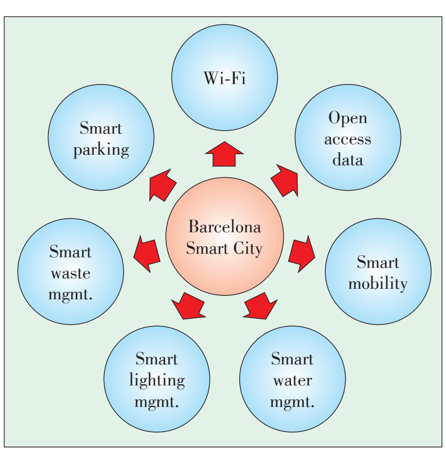
6 Internet of Things

6.1 Wi-Fi

Barcelona City Council aims to encourage citizens to access the internet and make it easier for citizens to incorporate this technology into their everyday lives. Mozilla Firefox, Microsoft Internet Explorer, and Google Chrome any browser can be used to access the internet through citizens' laptop, Smartphone, or computer. Wi-Fi service provided by Barcelona City Council enables citizens to connect to the Internet through Wi-Fi access points, in hotspots located several municipal amenities and various public access places. These places include centers for the elderly, civic centres, cultural centres and museums, sports centres, local authority and citizen advice offices, libraries, municipal markets, residential block interiors and enclosed parks with established opening times, night-time study rooms and neighbourhood centres. The government deployed its telecommunications network in 2007-12 with a wireless extension to reach any point of the public space. This network aims to provide services to the citizens and corporate workers. This network can be used as a backbone of a sensor and actuator network. This allows a better control of the city and the possibility to build applications [21]. Cisco is backbone network for City Access Network.

6.2 Open Access Data

This facility is free for smart phones including Android and iPhones. This allows all urban people to communicate with municipal mayors, concerned representatives and give their opinions, assess municipal rules, regulations and policies. This crowd sourcing became involved in the different participatory processes established in the city through on-line, on any topic that affects it. It could be bad situation about road, damage of street light, water leakage in particular place, accident, and in convenient law to citizens and so on. The app is designed to be very easy to use with rich GUI interface. The Open Data services collect all the public information from Barcelona's City Hall systems in Windows Azure SQL Database. The data includes street maps, details about public facilities, population, contractor profiles, city calendars, economy, businesses, travel and election results [22]. The huge amount of data gathered from different departments will be analyzed using Big Data Analytics software.



▲Figure 2. Barcelona Smart City dimensions.

6.3 Smart Mobility (e-vehicles)

Barcelona people really enjoy pleasant journeys with travel choices. In the trains, buses, city Wi-Fi connectivity is fully accessed. The city bus stops (**Fig. 3**) mounted with electronic displays or kiosks, give information automatically to passengers about buses arrival and departure timing. The touch screen facilities and Graphical User Interface facilities are really easy to operate. Manuel Sanroma, the Chief Information Officer (CIO), Barcelona City Council says that Smart bus stops change the typical experience of wasting passenger's time waiting for a bus. Payment of parking for cars, bikes on public places and road sides, will drastically bringing down the use of cars. This will indirectly reduce traffic jams in the city. Parking spaces are equipped with sensors and GIS integration leading to the commuter's easy way to park in free spaces with the help of Smart Mobile, PDAs. Bicycle, the last mile connectivity links different means of transportation stations and places. These bicycles will be available at all stations with annual usage payment. To date Barcelona [23], city has circa 500 hybrid taxis, 294 public electric vehicles, 130 electric motorbikes, an estimated 400 private electric vehicles on its streets, 262 recharging points. This leading public loveble journeys.

6.4 Smart Water Management System

The efficient consumption of water in cities is a basic ele-

▲Figure 3. Smart bus stop.

ment in sustainability programmers nowadays. The Smart Water Management System has become a key policy issue for the 21st century, as a growing number of factors are impacting the delivery of already scarce fresh water. Economic growth, seasonal climatic conditions and rising population are all mainly affecting availability of water resources. Moreover, a number of effects linked to climate change, such as lengthy droughts and extreme weather events, are worsening the situation [24]. In this light, Barcelona Smart City project is doing well for the city's green spaces and theme parks with smart water management techniques. The Smart Water Management system will optimize water consumption because it will irrigate with the proper amount of water according to weather conditions and plant needs. The four principles (4Rs: Reduce, Recycle, Reuse, Restore) are best practices of municipality for better water consumption. The Graphical User Interface (GUI) is user friendly. Sensors gather information about humidity, salinity, temperature, wind and several other factors that automatically regulate the amount of water by means of a program that can be managed with tubes, computers, smartphones, tablets and actuators. So far, nearly 77 fountains have been mounted in the entire city. District heating and cooling is one more kind of water technology in which two networks provide hot water in 64 buildings spanning an area of 21 km. Water theft and leakages can be tracked automatically. The Barcelona Smart City Deputy Mayor Antoni Vives, indicated that Barcelona is saving $58 million annually using Smart Water technology and stated that this new Smart Irrigation System in the city, will enable up to 25% saving of the water. So Smart Water Management System is efficient way to use the water in our daily life.

6.5 Smart Lighting System

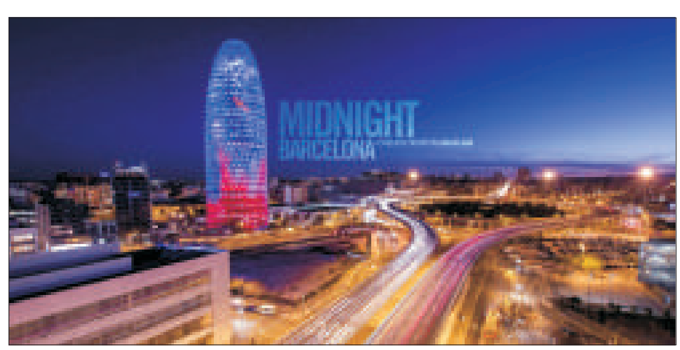
In the name of Smart Lighting System project, Barcelona developed a master plan in 2012. This project includes (1) remote control street level lighting, (2) transitioning 50 streets, and (3) connecting 1155 lamp posts to LED technology. From Smart Grid to self-sufficient blocks, Barcelona has developed a programme to achieve greater energy efficiency. We know that Smart Grid is a combination of information and communication applications that link generation, transmission, distribution, and customer end-use technologies. Internet of Things [25] has been boosted by Cisco in Barcelona, especially into more Smart Grid technologies. The city has deployed over 19,500 smart metres in the Olympic Villa. They are now extending the same project in city in conversion from Barcelona Brown Field City into Smart City. With Smart Grid technology, the city's entire power generation, transmission, distribution and consumption happens in efficient way. The power leakages will be detected and power theft be controlled. With its low power consumption LEDs, the Barcelona city night is shown in **Fig. 4**.

6.6 Smart Waste Management System

Barcelona´s Smart City project's Waste Management Systems is now a reality. Garbage vessels transmit signals to indicate they are over 80% full and should be emptied. Using Smart Mobile applications communication network, the signals are sent to a web-based software application used by the private MOBA's Smart Waste Management System. Sensor Technology 4.0., deliver a differentiated image of reality, and can transmit this image in real time via the web or internet. The garbage is collected separately in solid and liquid sorts with very high speed from homes and offices. In the software, the capacity of the container is visualized in a traffic light system, which is taken as the basis for planning the best route for waste collection-garbage trucks travel only to those containers that actually need to be emptied. Smart ultrasonic technology is combined with GSM communication technologies. Waste is recycled systematically and efficiently without environmental harm. Smart Waste Management involves [26] (1) regular reporting of measured fill levels and sensor data via mobile communication network, (2) robust ultrasonic sensor detects fill level regardless of the kind of waste, (3) fill level measurements as a basis for optimized routes for waste collection, and (4) reducing gas emissions and noise levels. These kinds of new waste management solutions even prevent the bad smells in residentail.

6.7 Smart Parking System

Barcelona Smart City includes around 500 parking wireless Fastprk sensors within Gran Via de Carles III, Av. Sarrià and Travessera de les Corts, at Les Corts District. This smart project uses Sigfox telecoms technology. It aims to reduce congestion in the area and to improve the drivers' experience while reducing CO_2 emissions. World sensing revolutionises traffic management and the industrial world through solutions based on wireless sensored networks allowing traffic detection and data capture in real time. The Smart Parking System based on electromagnetic sensors installed in each parking bay that senses the occupancy of each space, sending the information to a Central Management Unit (CMU). To deliver this information, Fastprk sensors are connected to the IoT network owned by Cellnex Telecom. This information is available in real time and is displayed through different panels in Pl. Neruda, Pl Hispanitat and Pl. Gaudí. Drivers will also be able to check this



▲**Figure 4. Smart lighting.**

data through a mobile phone apps, through a Web Application and on the website portal www.zonabus.cat [27]. Smart parking payment drastically reducing un necessary travellings.

7 Conclusions

For more than 2000 years, Barcelona has its own culture, customs, architecture, entertainment, business opportunities, and education facilities. Barcelona has its own vision and mission in which its architects, municipality representatives' and city dwellers jointly plan, design, invest in continuously for the development of city and for its citizen's services. In recent years, progress with Internet of Things technologies deployment in water, waste, gas and power kind of public utility services, made it number one Smart City in the world. Barcelona City Council encourages the use of mobile to access city services. The city has been pro-environment since the 1980s, and today it is a mature city concerned with environmental issues like waste, recycling, saving water and energy, and energy recovery. The striving for self-sufficient, with productive neighborhoods, living at a human speed and producing zero emissions motto and neighborhood has ensured quality of life to all citizens. Hence this city is a productive, open, inclusive and innovative city. This is the living city with enterprising people and healthy organized communities. Now it is the need of hour to construct such beautiful 100 more Barcelona Smart Cities, which are nothing but like Heavens on the Earth. With the help of technological God, IoT, anybody can reap all the civic services at any time, from any part of city, through any network via any device in these Heavens for Quality of Life.

References

[1] D. Kingsley, "Urbanisation in India—past and future," in *India's Urban Future*, R. Turner, Ed. Berkley, USA: University of California Press, 1962.
[2] P. Datta, "Urbanisation in India," in *European Population Conference*, Bratislava, Slovak Republic, Jun. 2006.
[3] E. B. Mijah, "Globalization, cities and the challenges of governance: a preliminary study of Kaduna City, Nigeria," *Journal of Law, Policy and Globalization*, vol. 30, pp. 61–67, 2014.
[4] E. Aibar and W. E. Bijker, "Constructing a city: the Cerdà plan for the extension of Barcelona," *Science, Technology & Human Values*, vol. 22, no. 1, pp. 3–30, 1997.
[5] International Telephone and Telegraph Corporation, *Barcelona; Barcelona, the Old City*. New York, USA: Bureau of Information Pro-España, 1927.
[6] C. Kennedy, J. Steinberger, B. Gasson, et al., "Greenhouse gas emissions from global cities," *Environmental Science & Technology*, vol. 43, no. 19, pp. 7297–7302, 2009.
[7] J. Laterasse, "The intelligent city," in *Telecom, Companies, Territories*, F. Rowe and P. Veltz Eds. Paris, France: Presses de L'ENPC, 1992.
[8] D. V. Gibson, G. Kozmetsky, and G., Smilor, R.W. Eds., *The Technopolis Phenomenon: Smart Cities, Fast Systems, Global Networks*. New York, USA: Rowman & Littlefield, 1992.
[9] W. J. Mitchell, *City of Bits*. Cambridge, USA: MIT Press, 1995.
[10] W. J. Mitchell, *E-topia: Urban Life, Jim—But Not As We Know It*. Cambridge, USA: MIT Press, 1999.
[11] B. Bowerman, J. Braverman, J. Taylor, et al., "The vision of a smart city," in *2nd International Life Extension Technology Workshop*, Paris, vol. 28, 2000.
[12] W. J. Mitchell, *ME++: The Cyborg Self and the Networked City*. Cambridge, USA: MIT Press, 2004.
[13] IBM. (2013, Apr. 2). *A Vision of Smarter Cities: How Cities Can Lead the Way into a Prosperous and Sustainable Future* [Online]. Available: http://www-03.ibm.com/press/attachments/IBV_Smarter_Cities_-_Final.pdf
[14] R. Giffinger, C. Fertner, H. Kramar, et al., *Smart Cities-Ranking of European Medium-Sized Cities*. Vienna, Austria: Vienna University of Technology, 2007.
[15] R. P. Dameri and A. Cocchia, "Smart city and digital city: twenty years of terminology evolution," in *X Conference of the Italian Chapter of AIS, ITAIS*, Milan, Italy, Dec. 2013.
[16] J. A. Benfield and W. J. Szlemko, "Internet-based data collection: Promises and realities," *Journal of Research Practice*, vol. 2, no. 2, Article D1, Jan. 2006.
[17] S. J. Best and B. S. Krueger, *Internet Data Collection*. Thousand Oaks, USA: Sage Publications, 2004.
[18] P. Baxter and S. Jack, "Qualitative case study methodology: Study design and implementation for novice researchers," Nova Southeastern University, Fort Lauderdale, USA, Qualitative rep. vol. 13 no. 4 pp. 544–559, Dec. 2008.
[19] MNT. (2014, May 16). *Barcelona is "iCapital" of Europe* [Online]. Available: http://europa.eu/rapid/press-release_IP-14-239_en.htm
[20] A. Steffen and A. Gore, *Worldchanging: A User's Guide for the 21st Century*. New York, USA: Harry N. Abrams, 2008.
[21] T. Gea, J. Paradells, M. Lamarca, and D. Roldan, "Smart cities as an application of internet of things: experiences and lessons learnt in Barcelona." in *Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS)*, Taichung, Taiwan, Jul. 2013. doi: 10.1109/IMIS.2013.158.
[22] MNT. (2015, Jun. 19). *City Deploys Big Data BI Solution to Improve Lives and Create a Smart-City Template* [Online]. Available: http://smartcitiescouncil.com/system/tdf/public_resources/Barcelona% 20deploys% 20big% 20data%20solution.pdf?file=1&type=node&id=712
[23] MNT. (2013, Dec. 15). *City Climate Leadership Awards* [Online]. Available: http://cityclimateleadershipawards.com/barcelona-barcelona-smart-city
[24] MNT. (2014, Dec. 26). *ICT as an Enabler for Smart Water Management* [Online]. Available: http://www.itu.int/dms_pub/itu - t/oth/23/01/T23010000100003PDFE.pdf
[25] S. Madakam, R. Ramaswamy, and S. Tripathi, "Internet of things (IoT): a literature review," *Journal of Computer and Communications*, vol. 3, no. 5, pp. 164–173, May 2015. doi: 10.4236/jcc.2015.35021.
[26] Pepperl-Fuchs. (2015, Oct. 20). *Smart Waste Management with Sensor Technology 4.0* [Online]. Available: http://www.pepperl-fuchs.com/global/en/27019.htm
[27] World Sensing. (2015, Apr.). *Worldsensing Traffic Providing Smart Parking Solution to Barcelona Touristic Coaches* [Online]. Available: http://www.worldsensing.com/news-press/worldsensing-traffic-providing-a-smart-parking-solution-to-barcelonas-touristic-coaches.html

Manuscript received: 2015-06-19

Biographies

Somayya Madakam (somu4smart@gmail.com) received his B.Tech (CSE) and MBA (IT) from Andhra University, India. He later worked for TISS, Mumbai for six years as a systems analyst cum programmer. Due to keen interest in teaching and research, he joined NITIE, Mumbai in 2012 to pursue the Fellow Program. His research topic is on "Internet of Things Technologies in Smart Cities: An Exploratory Study in India". During these three years, he presented and published circa 10 articles and posters in both national and International conferences. The article "Smart Cities: Six Dimensions" presented at ACIT-2014 conference, shot journal citations. The poster "The Lavasa Smart City is in Indian Clouds" got one of the best posters in ICCC-2015 conference. Besides, the article "A Re-View on Internet of Things" is presented in ICNGCCT-2014 conference. Further work includes Smart Homes (IEEE CPS), 100 New Smart Cities (NSITNSW 2015), and "GIFT Smart City: A Business Model" at (ICWR-2015).

Ramaswamy Ramachandran (ramaswamy2008@gmail.com) is a professor of National Institute of Industrial Engineering (NITIE), Mumbai with more than 30 years. He is expertise in the subject areas of Management Information System, Data Communications, Computer Networks, programming in C++, programming methodology, software engineering, IT strategy & Knowledge Management, Internet of Things and Smart Cities to name a few. He has vast experience in teaching, research and administrative work. He has made good corporate consultancy and written research papers, publications, and presentations in both national and international conferences. He bagged many prizes and awards during his rich academic periods including Best Professor too. He also trains the corporate people, public sector employees in Management Development Programmes (MDP) and Unit based Programmes (UBP).

DOI: 10.3969/j. issn. 1673-5188. 2015. 04. 002
http://www.cnki.net/kcms/detail/34.1294.TN.20151125.1735.004.html, published online November 25, 2015

Smart Cities in Europe and the ALMA Logistics Project

Didier El Baz[1] **and Julien Bourgeois**[2]
(1. CNRS, LAAS, 7 avenue du colonel Roche, F-31400 Toulouse, France,
Université de Toulouse, F-31400 Toulouse, France;
2. Université de Franche-Comté-FEMTO-ST Institute, UMR CNRS 6174, 1 cours Leprince-Ringuet, F-25200 Montbéliard, France)

Abstract
In this paper, a brief survey of smart citiy projects in Europe is presented. This survey shows the extent of transport and logistics in smart cities. We concentrate on a smart city project we have been working on that is related to A Logistic Mobile Application (ALMA). The application is based on Internet of Things and combines a communication infrastructure and a High Performance Computing infrastructure in order to deliver mobile logistic services with high quality of service and adaptation to the dynamic nature of logistic operations.

Keywords
smart cities; Internet of Things; logistics; combinatorial optimization; high performance computing

1 Introduction

The growth of cities has been particularly noticeable in the twentieth century and has raised many issues related to pollution, health, water distribution, logistics, and transport. The concept of smart cities has emerged recently as a way of addressing these issues using technology and social information. The European Union has promoted several smart cities projects with the goal of sustainable development. One of these projects is SmartSantander, a city-scale facility for experimental research on smart-city applications and services that are scalable, flexible, and open. The project involves the deployment of 20,000 sensors in several European cities, including Belgrade, Guildford, Lübeck, and Santander. IoT technologies and user acceptability will be the subjects of experimental research and testing.

Logistics and transport is of primary importance in a smart city. For logistics operators who deliver goods to customers, optimizing quality of service, e.g., ensuring on-time delivery for reasonable cost, is of major concern. This necessitates the optimization of truck loading and vehicle routing. The nature of logistics is dynamic—orders or cancellations may be made at any time, and transportation difficulties may arise at any time. These vicissitudes may be due to vehicle faults, traffic jams, or weather conditions.

In this paper, we concentrate on smart cities in Europe and present "A Logistic Mobile Application" (ALMA) project, which proposes a mobile, real-time, IoT-based approach to solving dynamic logistic problems and optimizing quality of service in logistics. Mobile devices like smart phones are used to report good delivery occurrences and incidents like an engine fault or a traffic jam; they are also used in order to launch computations related to the solution of a resulting routing problem on computing infrastructures in order to cope with incidents in real time. The ALMA project relies on a new high-performance computing (HPC) infrastructure that makes use of clusters, grids and volunteer computing, e.g., peer-to-peer networks via a broker that takes into account computational need and machines availability. The peer-to-peer concept has seen great developments with file sharing applications like Gnutella or FreeNet. Recent advances in microprocessors architectures, e.g., multicore processors and advances in high bandwidth networks permit one to consider high performance volunteer computing as an economic and attractive solution. The ALMA project relies also on new optimization algorithms for the solution of combined truck loading and vehicle routing problems.

In section 2, we present a brief overview of smart city projects in Europe. Section 3 deals with logistics issues. We present ALMA architecture in section 4; in particular, we detail the communication infrastructure and the HPC infrastructure. Some preliminary computational results are presented in section 5. Finally, conclusions and future work are introduced in section 6.

2 Smart Cities

From the time "smart cities" was first coined in 2000 [1], there have been numerous definitions of what a smart city is [2]–[6]. In [1], a smart city is "the urban center of the future, made safe, secure environmentally green, and efficient because all structures—whether for power, water, transportation, etc. are designed, constructed, and maintained making use of advanced, integrated materials, sensors, electronics, and networks which are interfaced with computerized systems comprised of databases, tracking, and decision-making algorithms." In this very first definition, the technological part is emphasized but the citizen are forgotten which is corrected in this later definition from [4]: "Smarter Cities are urban areas that exploit operational data, such as that arising from traffic congestion, power consumption statistics, and public safety events, to optimize the operation of city services. The foundational concepts are instrumented, interconnected, and intelligent. This approach enables the adaptation of city services to the behavior of the inhabitants, which permits the optimal use of the available physical infrastructure and resources."

The turning point in Europe for the definition of smart cities is a report of the Centre of Regional Science at Vienna University of Technology [3], which identifies six main axes defining a smart city. These axes are: smart governance (participation), smart mobility (transport and information and communication technologies, ICT), smart environment (natural resources), smart people (social and human capital), smart living (quality of life) and a smart economy (competitiveness). The smart city is also defined as "A city well performing in a forward-looking way in economy, people, governance, mobility, environment, and living, built on the smart combination of endowments and activities of self-decisive, independent and aware citizens."

In this report, a definition of the smart city within Europe emerged [7]: a smart city "is a city seeking to address public issues via ICT-based solutions on the basis of a multi-stakeholder, municipally based partnership". This definition is still contested but is generally accepted as the official definition.

To be classified as a smart city, a city must contain at least one initiative that addresses one or more of the following characteristics: smart governance, smart people, smart living, smart mobility, smart economy, and smart environment. ICT initiatives based on these characteristics aim to connect existing and improved infrastructure to enhance the services available to stakeholders (citizens, businesses, communities) in a city.

2.1 Development of Smart Cities in Europe

If this latter definition is used, it is possible to evaluate the status of smart city within the EU, counting only its 28 member states. Ninety percent of cities with a population of more than 500,000 have implemented or are in the process of implementing smart cities initiatives. This percentage drops to 51% for cities with a population of more than 100,000. This means that the concept of smart city is well-known in EU. The champions are the UK, Spain, Italy, Austria, Denmark, Norway, Sweden, Estonia and Slovenia.

2.2 State of the Art

Smart Urban Spaces (SUS) [8] is a project funded by EU in 2009. The aim of this project was to deploy innovative mobile services in real conditions using a network of European cities. Different applications have been developed like ticketing service but the most interesting application is a museum quest a quiz installed at the Caen museum. This application uses near-field communication (NFC) tags to ask questions about the item displayed.

The main concern at EU level for a smart city is energy efficiency. Many different EU projects have been funded to study how to enhance energy usage in future cities. Building Energy Decision Support Systems for Smart Cities (BESOS) [9] integrates diverse and heterogeneous energy-management systems into a single platform, enabling higher-level applications to take care of data and services from multiple sources. Better energy efficiency in buildings is also the objective of the Control and Optimisation for Energy Positive Neighbourhoods (COOPERATE) [10] project, which has the same idea of offering a single interface for many different sensors and data. Decision Support Advisor for Innovative Business Models and Use Engagement for Smart Energy Efficient Districts (DAREED) offers approximatively the same service but at a wider range. It also puts the citizen at the center of the system, providing information and action that can be taken to reduce energy consumption. Within the same scope one can cite District of the Future (DoF) [11] and Energy Efficiency in the Supply Chain through Collaboration, Advanced Decision Support and Automatic Sensing (e-SAVE).

Other projects, such as Energy Forecasting (NRG4CAST), focus on efficient energy distribution in urban and rural communities through real-time management, analytics and forecasting. The Energy Positive Neighbourhoods Infrastructure Middleware based on Energy-Hub Concept (EPIC-HUB) project developed a middleware to ease this task.

Finally, keeping the ease of use in mind the Environmental Services Infrastructure with Ontologies (ENVISION) project aims to help non-ICT specialists discover and combine environmental services.

In the smart governance item, the flagship is Helsinki with the Infoshare project [12]. Infoshare gives free access to various urban statistics which can be used by businesses, academia and research institutes, governmental institutes or citizens. These data are covering many different aspects of Smart governance like living conditions, employment, transport, economics and so on.

3 Logistics

Logistic applications involve difficult problems, most of which are NP-complete problems [13]–[17]. The ALMA logistic application considers combined truck loading and vehicle routing problems. Treatment of vehicle routing problems in conjunction with truck loading is very attractive in just-in-time distribution context. Indeed the stock can be close to zero.

This technique is used more and more in car manufacturing and mass-market retailing. Despite the advantages of just-in-time distribution, in particular, cost reduction, this technique may create weaknesses in the logistic chain in case of failures. Therefore, it is necessary to treat dynamically and as quickly as possible the events that may perturb the correct working of the logistic chains.

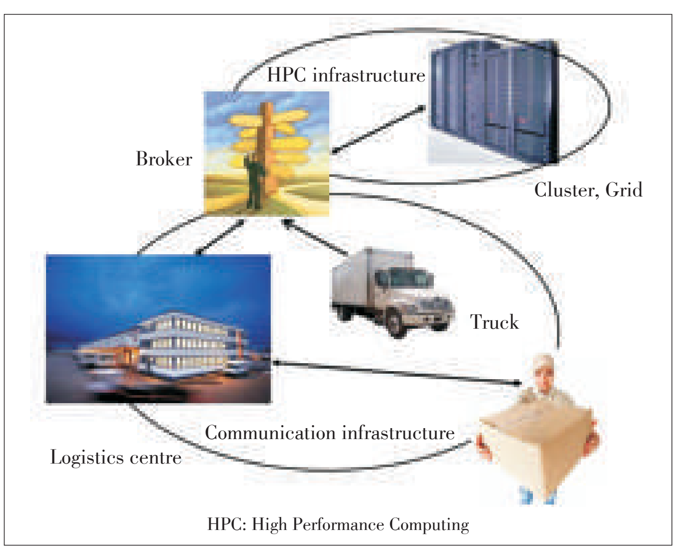
Treatment of vehicle routing problems in conjunction with truck loading has been discussed in the literature [13]–[16]. The ALMA logistic application concentrates on dynamic logistic problems whereby dynamism results from new orders, cancellations, as well as traffic incidents that may occur at any time. This leads to extremely difficult problems. Our approach is based on the approximate solution of truck loading problems via strip generation and beam search [17]–[19]. Vehicle-routing problems are solved via Ant Colony Optimization (ACO) [20]. This approach relies on parallel and distributed computing systems because those optimization problems are difficult to solve. We consider clusters, volunteer computing and peer-to-peer infrastructures.

4 Global Alma Architecture

The ALMA logistic application relies on two infrastructures: a communication infrastructure and an HPC infrastructure. **Fig. 1** displays the infrastructures of the mobile application ALMA [21].

4.1 The Communication Infrastructure

Goods to be delivered are identified by tags. When a good is delivered, the transporter scans the tag and transmits the information in real time to the logistics centre with a smart phone connected to the Internet 3G. The mobile application is based on the existing telecommunication infrastructure. Similarly, the transporter informs the center in real time of traffic incidents, like road closed and traffic jam. In case of problems, e.g. traffic incidents, the proposed initial route may not be valid. Thus the transporter uses also the mobile application to ask for a new route. The request for a new route is transmitted to the broker of the HPC infrastructure.



▲**Figure 1. Communication and HPC infrastructures of the mobile application ALMA.**

4.2 The HPC Infrastructure

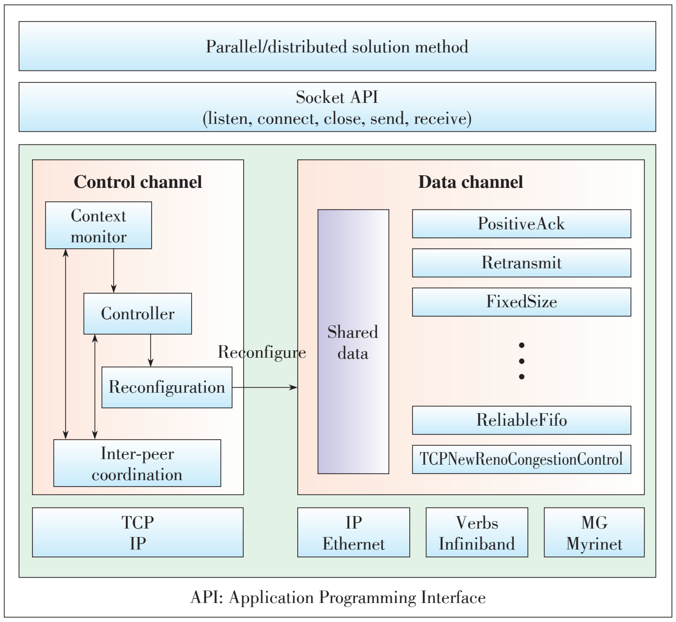
4.2.1 The Broker

The broker is designed in order to select a convenient HPC infrastructure from several available parallel or distributed computing systems. These systems may be clusters or peer-to-peer networks. For a given vehicle-routing problem and method, the broker selects a convenient topology and number of machines. This represents an evolution from the approach in [22]. The main goal of the broker is to select a computing infrastructure that satisfies the real-time constraints of the application. The requests of vehicle routing solution are associated with a deadline for result reception in order to limit important vehicle immobilization and blocking of the logistics application. The selection of unsuitable infrastructure leads to a suboptimal solution.

Two main phases are considered for brokering: first, the supervision of available resources, e.g. clusters or peer-to-peer networks. Secondly, the prediction of computation time for the considered problem and selected method. We note that these steps can be iterated several times in order to improve prediction. Reference is made to [23] to [25] for previous work on performance prediction of HPC applications on distributed computing infrastructures.

4.2.2 The Environment for Computing

The environment for computing is an extension of peer-to-peer distributed computing (P2PDC) [22]. P2PDC is a decentralized environment for peer-to-peer high-performance computing. P2PDC is a multinetwork environment that supports Infiniband, Myrinet and Ethernet networks. P2PDC is particularly used to task parallel applications. It is intended for scientists who want to solve difficult optimization problems or numerical simulation problems via distributed iterative methods that lead to frequent direct data exchanges between peers. References [26] and [27] provide more details and extensions of P2PDC. P2PDC relies on the use of the P2PSAP self-adaptive communication protocol [28] (**Fig. 2**) and a reduced set of communication operations, i.e., P2Psend, P2Preceive and P2Pwait in order to facilitate programming. The programmer cares only about the choice of distributed iterative scheme of computation, e.g., synchronous or asynchronous, that needs to be implemented and does not care about the communication mode between any two nodes. The programmer can also select a hybrid iterative scheme of computation, whereby computations are locally synchronous and asynchronous at the global level.



▲**Figure 2. P2PSAP protocol architecture.**

P2PSAP dynamically chooses the most appropriate communication mode between any two peers according to a decision taken at application level, such as scheme of computation and elements of context like network topology at transport level. In the hybrid case, the communication mode between peers in a group of nodes that are close and that present the same characteristics is synchronous, and the communication mode between peers in different groups is asynchronous. The decentralized environment of P2PDC is based on a hybrid topology manager and a hierarchical task-allocation mechanism which make P2PDC scalable. P2PSAP communication protocol was designed first as an extension of the CTP transport protocol [29] based on the CACTUS framework, which uses microprotocols [30].

The CTP protocol includes a wide range of micro-protocols including a small set of basic micro-protocols like Transport Driver, Fixed Size or Resize and Checksum that are needed in every configuration and a set of micro-protocols implementing various transport properties like acknowledgements, retransmissions, error correction and congestion control. The P2PSAP communication protocol takes into account Ethernet, Infiniband and Myrinet clusters. Reference is also made to [31] for details on peer-to-peer computing.

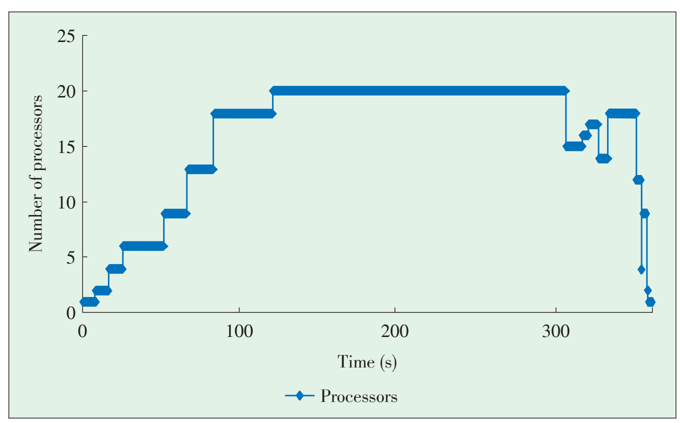
5 Experimental Results

Here, we consider loading problems and present preliminary experimental results obtained for a 2D cutting stock problem solved using a two-stage, two-dimensional method based on strip generation and beam search via the decentralized environment P2PDC on the Grid 5000 testbed. For details on the two-stage two-dimensional method based on strip generation and beam search see [32] and [33].

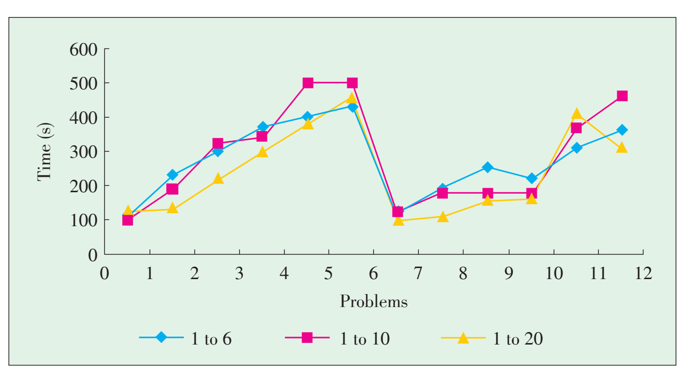
Fig. 3 shows the number of active processors during the solution of a cutting stock problem in function of the time. A maximum of twenty processors were allocated to this particular problem. The number of active processors varies according to the evolution of the algorithm, i.e., the need of computing resources to treat the problem in parallel. In the beginning, the solution requires few computing resources because the number of nodes to explore is small. The number of processors increases with time because more and more nodes to explore are created until the limit is met, i.e., the maximum number of twenty processors that were allocated to the solution of this problem. At the end of this solution, the number of active processors decreases because the number of nodes to explore decreases.

Obtaining a good approximation of the best solution at a given processor and communicating it to other processors means that the need for computing resources can sometimes be significantly decreased. This is what we observe when the number of processors decreases suddenly from twenty to fifteen. Nevertheless, we observe that the number of computing resources required may increase for a while before finally tending to zero at the end of the computation.

Fig. 4 displays solution times for several instances of cutting stock problems according to the maximum number of allocated processors.



▲**Figure 3. Cutting stock problem: number of active machines.**



▲**Figure 4. Cutting stock problem: solution time according to the maximum number of allocated machines.**

Three cases are considered: a case with a maximum number of six computing nodes (diamonds), a case with maximum number of ten nodes (squares), and a case with twenty nodes (triangles). In general, the more processors that are allocated, the smaller the solution time. This shows that our approach is scalable in terms of the number of computing resources, i.e., the number of processors in the computing system. The design of the architecture of the HPC infrastructure also makes our approach scalable when the complexity of the problem increases, i.e., when the number of goods and vehicles increases or the size of the city/conurbation increases due to the dedicated brokering system and large number of computing resources available via cluster or volunteer computing systems.

References [34] and [35] give details on peer-to-peer distributed algorithms for 2D Cutting stock problems. Reference [36] describes distributed branch and bound on peer-to-peer networks.

6 Conclusions

In this paper, we have presented an overview of smart city projects in Europe. We have shown that transport and logistics projects are prominent in smart cities. We have detailed the logistics mobile application ALMA that is based on the Internet of Things. ALMA addresses dynamic logistics problems whereby new orders or cancellations or traffic incidents may occur at any time. The ALMA application permits one to communicate in real time the information regarding delivery of goods.

The logistics application ALMA combines a communication infrastructure and a parallel/distributed computing infrastructure in order to obtain rapidly new routes for transporters that deliver goods to customers in case of incidents like traffic jam. The HPC infrastructure makes use of a broker to select the convenient parallel/distributed computing system as well as the number of computing nodes to perform computations according to a fixed deadline. Clusters or peer-to-peer infrastructures can be selected from a pool of available parallel/distributed computing systems. The computing infrastructure makes use of the high - performance computing decentralized environment P2PDC.

The mobile application ALMA also addresses combined truck loading and vehicle routing problems that lead to very complex optimization problems. Preliminary computational results for cutting stock problems solved on Grid 5000 have been presented and analyzed in the paper. This permits us to illustrate the interest of the proposed approach.

We are presently extending the P2PSAP communication protocol and P2PDC decentralized environment to multiple network context, i.e., Infiniband, Myrinet and Ethernet networks, and heterogeneous architectures combining multicore CPUs and GPUs. The self-organizing strategies are also studied for deployment and efficiency purposes or for insuring everlastingness of applications in hazardous situations or in the presence of faults.

Acknowledgment

The authors wish to thank Agence Nationale de la Recherche (ANR) for support through several contracts. Experiments presented in this paper were carried out using the Grid'5000 experimental testbed being developed under the INRIA ALADDIN development action with support from CNRS, RENATER and several Universities as well as other funding bodies.

References

[1] B. Bowerman, J. Braverman, J. Taylor, H. Todosow, and U. Von Wimmersperg, "The vision of a smart city," In *2nd International Life Extension Technology Workshop*, Paris, France, Sept. 2000.

[2] A. Caragliu, C. Del Bo, and P. Nijkamp, "Smart cities in Europe," *Journal of Urban Technology*, vol. 18, no. 2, pp. 65–82, 2011.

[3] R. Giffinger, C. Fertner, H. Kramar, *et al.*, "Smart cities-ranking of European medium-sized cities," Vienna University of Technology, Vienna, Austria, Tech. rep., 2007.

[4] C. Harrison, B. Eckman, R. Hamilton, *et al.*, "Foundations for smarter cities," *IBM Journal of Research and Development*, vol. 54, no. 4, pp. 1–16, Jul. 2010.

[5] T. Nam and T. A. Pardo, "Conceptualizing smart city with dimensions of technology, people, and institutions," in *Proc. 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times*, New York, USA, Jun. 2011, pp. 282–291.

[6] K. Su, J. Li, and H. Fu, "Smart city and the applications," in *International Conference on Electronics, Communications and Control*, Ningbo, China, Sept. 2011, pp. 1028–1031. doi: 10.1109/ICECC.2011.6066743.

[7] C. Manville, G. Cochrane, J. Cave, *et al.*, "Mapping smart cities in the EU," European Parliament, Brussel, Belgium, Rep. IP/A/ITRE/ST/2013-02 PE 507.480, 2014.

[8] S. Chaumette and J. Ouoba, "A multilevel platform for secure communications in a fleet of mobile phones," in *6th International Conference on Mobile Computing, Applications and Services*, Austin, USA, Nov. 2014, pp. 173–174.

[9] W. Apolinarski, "Establishing secure intelligent environments," in *Workshop Proc. 11th International Conference on Intelligent Environments*, Prague, Czech Republic, Jul. 2015, pp. 43–45.

[10] T. Greifenberg, M. Look, and B. Rumpe, "Integrating heterogeneous building and periphery data models at the district level: the NIM approach," in *Proc. 10th ECPPM eWork and eBusiness in Architecture, Engineering and Construction*, Vienna, Austria, Sept. 2014, pp. 821–828.

[11] District of Future. (2015, Feb. 25–26). "Europe: A better place to live and work," *Sabadell Smart Congress 2015* [Online]. Available: http://www.districtoffuture.eu/index.php/mod.eventos/mem.detalle/id.22/relcategoria.104

[12] *Helsinki Region Infoshare* [Online]. Available: http://www.hri.fi/en

[13] M. Iori, J. J. Salazar Gonzalez, and D. Vigo, "An exact approach for the symmetric capacited vehicule routing problem with two dimensional loading constraints," DEIS, University of Bologna, Bologna, Italy, Tech. Rep. OR/03/04, 2003.

[14] M. Gendreau, M. Iori, G. Laporte, and S. Martello, "A Tabu search heuristic for the vehicle routing problem with two-dimensional loading constraint," *Management Science*, vol. 40, no. 10, pp. 1276–1290, 1994.

[15] E. Choi and D. W. Tcha, "A column generation approach to the heterogeneous fleet vehicle routing problem," *Computers and Operations Research*, vol. 34, no. 7, pp. 2080–2095, Jul. 2007.

[16] K. Ganesh and T. T. Narendran, "CLOVES: a cluster-and-search heuristic to solve the vehicle routing problem with delivery and pick-up," *European Journal of Operational Research*, vol. 178, no. 3, pp. 699–717, May 2007.

[17] M. Hifi, R. M'Hallah and T. Saadi, "Approximate and exact algorithms for the double-constrained two-dimensional guillotine cutting stock problem," *Computational Optimization and Applications*, vol. 42, no. 2, pp. 303–326, Mar. 2009. doi: 10.1007/s10589-007-9081-5.

[18] M. Hifi, R. MHallah, and T. Saadi, "Algorithms for the constrained two-staged two-dimensional cutting problem," *INFORMS Journal on Computing*, vol.20, no. 2, pp. 212–221, 2008. doi: 10.1287/ijoc.1070.0233.

[19] M. Hifi and T. Saadi, "A cooperative algorithm for constrained two-staged two-dimensional cutting problems," *International Journal of Operational Research*, vol. 9, no.1, pp. 104–124, 2010.

[20] M. Dorigo, "Optimization, learning and natural algorithms", PhD thesis, Po-

litecnico di Milano, Italy, 1992.
[21] D. El Baz, J. Bourgeois, T. Saadi, and A. Bassi, "ALMA, a logistic mobile application based on the internet of things," in *2013 IEEE International Conference on Internet of Things*, Beijing, China, Aug. 2013, pp. 355–358. doi: 10.1109/GreenCom-iThings-CPSCom.2013.78.
[22] T. T. Nguyen, D. El Baz, P. Spiteri, J. Jourjon, and M. Chau, "High performance peer-to-peer distributed computing with application to obstacle problem," in *Proc. 24th IEEE IPDPSW*, Atlanta, USA, May 2010, pp. 1453–1461. doi: 10.1109/IPDPS.2010.5470930.
[23] B. Cornea, J. Bourgeois, T. T. Nguyen, and D. El Baz, "Performance prediction in a decentralized environment for peer to peer computing," in *Proc. 25th IEEE IPDPSW*, Anchorage, USA, May 2011, pp. 1613–1621. doi: 10.1109/IPDPS.2011.321.
[24] B. Cornea and J. Bourgeois, "Performance prediction of distributed applications using block benchmarking methods," in *Proc. 19th Conference on Parallel, Distributed and Networked-Based Processing*, Ayia, Napa Cyprus, Feb. 2011, pp. 183–190.
[25] B. Cornea and J. Bourgeois, "A framework for efficient performance prediction of distributed applications in heterogeneous systems," *The Journal on Supercomputing*, vol. 62, no. 3, pp. 1609–1634, Dec. 2012. doi: 10.1007/s11227-012-0823-5.
[26] B. Cornea, J. Bourgeois, T. T. Nguyen, and D. El Baz, "Scalable performance predictions of distributed peer-to-peer application," in *Proc. 14th IEEE International Conference on High Performance Computing and Communication*, Liverpool, UK, 2012, pp. 193–201. doi: 10.1109/HPCC.2012.34.
[27] T. Garcia, M. Chau, T. T. Nguyen, D. El Baz, and P. Spiteri, "Asynchronous peer-to-peer distributed computing for financial applications," in *Proc. 25th IEEE IPDPSW*, Anchorage, USA, May 2011, pp. 1453–1461. doi: 10.1109/IPDPS.2011.292.
[28] D. El Baz and T. T. Nguyen, "A self-adaptive communication protocol with application to high performance peer-to-peer distributed computing," in *Proc. 18th Conference on Parallel, Distributed and Networked-Based Processing*, Pisa, Italy, Feb. 2010, pp. 327–333.
[29] G. T. Wong, M. A. Hiltunen, and R. D. Schlichting, "A configurable and extensible transport protocol," in *Proc. IEEE INFOCOM' 01*, Anchorage, USA, 2001, pp. 319–328. doi: 10.1109/INFCOM.2001.916714.
[30] M. A. Hiltunen, "The Cactus approach to building configurable middleware services," in *Proc. DSMGC2000*, Nuremberg, Germany, 2000.
[31] D. El Baz and G. Jourjon, "Some solutions for Peer to Peer Global Computing," in *Proc. 13th Conference on Parallel, Distributed and Networked-Based Processing*, Lugano, Switzerland, Feb. 2005, pp. 49–58.
[32] M. Hifi and R. M'Hallah, "An exact algorithm for constrained two-dimensional two-staged cutting stock problems," *Operations Research*, vol. 53, no. 1, pp. 140–150, 2005.
[33] M. Hifi and T. Saadi, "A parallel algorithm for constrained two-staged two-dimensional cutting problems," *Computers and Industrial Engineering*, Vol. 62, No. 1, pp. 177–189, 2012.
[34] M. Hifi, T. Saadi, AND N. Haddadou, "High performance peer-to-peer distributed computing with application to constrained two-dimensional guillotine cutting problem," in *Proc. 19th Conference on Parallel, Distributed and Networked-Based Processing*, Ayia, Napa Cyprus, Feb. 2011, pp. 552–559.
[35] D. El Baz, M. Hifi, and T. Saadi, "Peer-to-peer solution of 2D-cutting stock problems," in *Proc. 11th Cologne-Twente Workshop on Graphs and Combinatorial Optimization*, Munich, Germany, May 2012, pp. 116–120.
[36] M. Djamaï, B. Derbel, and N. Melab, "Distributed B&B: a pure peer-to-peer approach", in *Proc. 25th IEEE IPDPSW*, Anchorage, USA, May 2011, pp. 1788–1797. doi: 10.1109/IPDPS.2011.337.

Manuscript received: 2015-08-10

Biographies

Didier El Baz (elbaz@laas.fr) received his Dr. Engineer degree in Electrical Engineering and Computer Science from INSA Toulouse France in 1984 and was a visiting scientist in the Laboratory for Information and Decision Systems, MIT, USA in 1984-1985. Dr. El Baz received the Habilitation à Diriger des Recherches (HDR) from INP Toulouse in 1998. His fields of interest are in parallel and distributed computing, combinatorial optimization and IoT. Dr. El Baz has co-authored more than 120 international publications. He was the program chair of International Conference on Parallel, Distributed and Networked-Based Processing (PDP) 2008 and PDP 2009. He was also the program chair of IEEE CSE 2014, the general chair of IEEE CSE 2015, the executive chair of IEEE ScalCom 2015, the general chair of IEEE ScalCom 2016, IEEE UIC 2016, and IEEE ATC 2016, and the general co-chair of IEEE iThings 2013.

Julien Bourgeois (Julien.Bourgeois@univ-fcomte.fr) is a professor of computer science at the University of Franche-Comté (UFC) in France. He is part of the FEMTO-ST institute (UMR CNRS 6174) where he is leading the complex networks team. His research interests are in distributed intelligent MEMS (DiMEMS), P2P networks and security management for complex networks. He has been invited professor at Carnegie Mellon University (US) from 9/2012 to 8/2013, at Emory University (US) in 2011 and in Hong Kong Polytechnic University in 2010, 2011 and 2015. He led different funded research projects (Smart Surface, Smart Blocks, Computation and coordination for DiMEMS). He has worked for more than 10 years on these topics and has co-authored more than 120 international publications and communications. He has served as PC members and chaired various conferences (IEEE iThings, IEEE HPCC, Euromicro PDP IEEE GreenCom , IEEE CPSCom, GPC, etc. Apart from its research activities, he is acting as a consultant for the French government and for companies.

DOI: 10.3969/j. issn. 1673-5188. 2015. 04. 003
http://www.cnki.net/kcms/detail/34.1294.TN.20151118.1640.004.html, published online November 18, 2015

Smart City: On Urban Operational Collaboration

Rui Cao and Weidong Kou
(IBM, Beijing 100101, China)

Abstract

This paper expounds the origin of urban operational coordination problem in historical setting, points out that operational coordination problem is essential to cities, induces the major challenges and opportunities for urban operating coordination at present, and takes IBM Intelligent Operation Center as example to illustrate the typical solutions with the detailed case study of the Intelligent Operation Center in Rio de Janerio.

Keywords

smart city; urban operation; coordination; Intelligent Operation Center (IOC)

1 Introduction

Since the "smart city" concept was introduced by IBM in 2008 [1], it has been the subject of much discussion in terms of theories, standards, and solutions. It has been explored on a practical level across the globe. However, the city as a human society, the most complex and sophisticated system, issued as culture of human civilization [2], how to efficiently operate it is the problem that should be fully taken into account at the beginning of urban construction. It has been found out that the origin of inefficiency of urban construction in operational aspects comes from the repeated constructions and conflicts between different parties.

This paper will discuss smart city in terms of urban operational coordination, on its brief history, features and best practices nowadays, and also probe into the future about urban operational coordination.

2 Origin of the Urban Operational Coordination Problem

A city evolves from the countryside, and it is a more complex form for the gregarious than the countryside. Cities originated between BC 3500 to BC 3000 in Mesopotamia [3]. In the 5,000 years prior to the Industrial Revolution, cities evolved slowly. In ancient cities with limited scale and simple functions, urban operation is simple and pure: on Sundays, people go to church with their families when they heard the bow bell, or gather around in the citizen square to vote or discuss.

As cities have expanded and become more and more complex, with various functions, urban operation has become a problem, and urban operational coordination problem has become crucial.

2.1 Changes of Urban Scale

Tacitly, urban scale was considered to be within a day's walking distance from anywhere of the city both in ancient western and eastern culture. Citizens could walk to school, to the shops, to hospital, and of course, to the square. In the Republic, Plato believed that an ideal city should be the size of a speaker voice can be heard. And ancient Chinese describe city scale as "three-mile inner city, seven-mile outer city". [4]

In the Middle Ages, the London expanded as far as the reach of bow bell of St Mary [5]. Xi'an was the capital of 13 dynasties in Chinese history, and it used to be one of the most prosperous cities in the world. The perimeter of Xi'an Circumvallation, which was built around AD600, is 13.74 km. Each wall extends between 2.6 km to 4.2 km. [6]

Nowadays, over 50% of the world's population live in a city. That is equivalent to over 3 billion people [7]. London covers 1577.3 km^2 and has more than 8 million residents [8]. The perimeter of Beijing's 5th Ring is around 100 kms [9], and it is always the case in rush hours that you have to drive 2 to 3 hours in order to across the city from east to the west; more than 20 million people live in this mega city.

Not any square can accommodate all citizens, and not any voice can be heard at the same time. The tremendous change on scale is not simply quantitative change; the concept of modern city differs a lot from that in the old times.

2.2 The Diversification of Urban Functions and the Division of Functional Departments

Cities originate from religious activities [10]. People gather around at a certain place to worship their ancestors, pray, and trail sometimes. The initial function of city is a religious center. Authority always accompanies religion, especially in the early stage of human history. Not surprisingly, the city became a center of authority. To satisfy the demands of rulers, monks and nobility, more and more people come to the city center. They work there, trade there, and entertain there. Naturally, the city became trade center, traffic center, and gradually culture center, art center, economic center, etc. The city is like magnet in the way it attracts people and produces a kind of chemical reaction. It creates numerous possibilities between groups of people.

To regulate the "product" of the chemical reactions, numerous departments have been set up—from water management to public safety, from traffic department to central government. Recently a report revealed that China has around 40 million public service staff and 1.3 million agencies while some of them have more than one affiliates. [11]

3 Challenges and Opportunities of Urban Operational Coordination in Modern City

In the present day, most cities across the globe are facing a lot of operational issues. This in turn is deteriorating the quality of services that are being delivered to their citizens. To ensure safety and provide basic utilities, public transport, infrastructure facilities, so on, cities need to collate huge amounts of information from diverse sources and at the same time facilitate real time communication and collaboration among various city agencies. However, there are some big challenges.

3.1 Challenge: Information Islands

There are two origins of Information Island: collaboration mechanism and information system.

The water department plans to work on a city street in June, while in July, the utilities team plans to replace a gas line in the same location. How many times would the road be dug? It depends on whether the information has been shared between the two bureaus. This kind of information islands is caused by collaboration mechanism.

Learned from experiences in IBM smart city projects related to urban information system, the systems of different parties rarely compatible with each other. The way critical information is often stored hinders situational awareness and makes it difficult for various departments to coordinate emergency response efforts:

- Critical information is often stored in multiple disparate systems, across multiple, disconnected departments, hindering situational awareness and making it difficult for city officials to coordinate agency efforts.
- They lack a single, integrated view of events, incidents or impending crises, and the ability to rapidly share information.

Without a single, integrated view of events, incidents or impending crises, and without the ability to rapidly share information, a city might be unable to deliver services in a sustainable way, protect citizens, or drive economic growth for the future. However, it is not fair to criticize urban departments and their lack of top-level system design because the systems were built as cities were developing. It is a progressive process, but we have suffered too much. It becomes the bottleneck of urban operation capability.

3.2 Challenge: Limitation of Urban Management Capability

As mentioned in section 2.2, cities have various functions with considerable numbers of organizations supporting them. Collaboration across these organizations is critical for addressing crises, completing projects, and increasing the efficiency of daily operations. However, coordination different domains is not easy. The cross-domain collaboration capability gradually becomes bottleneck of urban management.

According to the 1907 Survey of New York traffic, at that time, the carriage moved at an average speed of 11.5 mph. In the 1960s, for a car slowly driving on the road, the average speed was about six miles per hour [12]. This may precisely demonstrate what happened without proper management in the 1960s. In the 21st century, although people have much faster cars and plenty of advanced technology to promote traffic management, the average driving speed is around 13 mph on the street on weekdays [13].

Transportation is just part of the problem. Carrying capacity and management capability of existing systems are also constantly being challenged by the rapid spread of infectious diseases—the Korean MERS [14] virus carrier concealed his illness when traveling to Hong Kong, Shenzhen and Huizhou and put millions of people in danger of being infected—by vicious terrorist group event—boomers set off booms on 2013 Boston Marathon [15], caused four died and hundreds of injury. There are still lots of challenges out there, more complex and sophisticated than ever.

3.3 Opportunity: Advanced Information Technology

The past two decades have seen rapid advances in sensors, database technologies, search engines, data mining, machine learning, statistics, distributed computing, visualization, and modeling and simulation. These technologies, which collectively underpin big data, are allowing organizations to acquire, transmit, store, and analyze all manner of data in greater volume, with greater velocity. In terms of the individual, internet and mobile/wearable devices enable people to continuously obtain or create data, location information, social opinion, physical data, etc. The increasing volume and detail of information captured by enterprises, the rise of multimedia, social media, and the Internet of Things will fuel the exponential growth of data for the foreseeable future.

In applying advanced information technologies in urban operation, sensors will reduce or eliminate traffic jams and optimize energy demand. New mapping technologies will identify and help to correct urban problems that no one could see before. Social media and crowdsourcing will target big and small problems, down to which potholes to fix—maybe even make governments work again. Wireless devices and networks make up the evolution of the digital nervous system (all-seeing) "eyes" and (all-hearing) "ears" [16]. Considering the city as a huge nervous system, even tiniest stimulation can be perceived. The challenge is to build on this and other successes with a new generation of data tools that help us identify and collaborate effectively, especially on urban challenges.

3.4 Opportunity: Subversion of Coordination Pattern

As is mentioned in section 3.3, information technology enables non-distance, non-time equation and non-cost communication among online people and things. Ideas spread with no limitations; real-time gathering and sharing of knowledge is easy and almost free. Traditional management and collaboration patterns are being subverted.

The winning strategy of DARPA Network Challenge perfectly illustrates how people leverage internet to efficiently collaborate. In the challenge, teams had to find 10 red weather balloons deployed at undisclosed locations across the continental United States. The first team to correctly identify the locations of all 10 won a $40,000 prize. The winning MIT team had more than 5000 participants, all of whom were leveraging the Internet. Their strategy was to recruit participants, and the prize money was distributed up the chain of participants leading to successful balloon spotting. All prize income remaining after distribution to participants to be given to charity. The team only began with four initial participants. And finally, this team found 10 red balloons across the US within 7 hours [17].

There are also many other cases that show the power of information technology in connecting people to achieve something humans never could have imagined decades ago. A classic case in politics is how the Arab Spring revolution be affected by Facebook [18].

4 Solutions and Practice on Addressing Urban Operational Coordination Problem at Present

At present, there are some solutions and practices related to urban operational coordination issue. IBM Intelligent Operation Center (IOC) is a typical and relatively mature solution in the market [19].

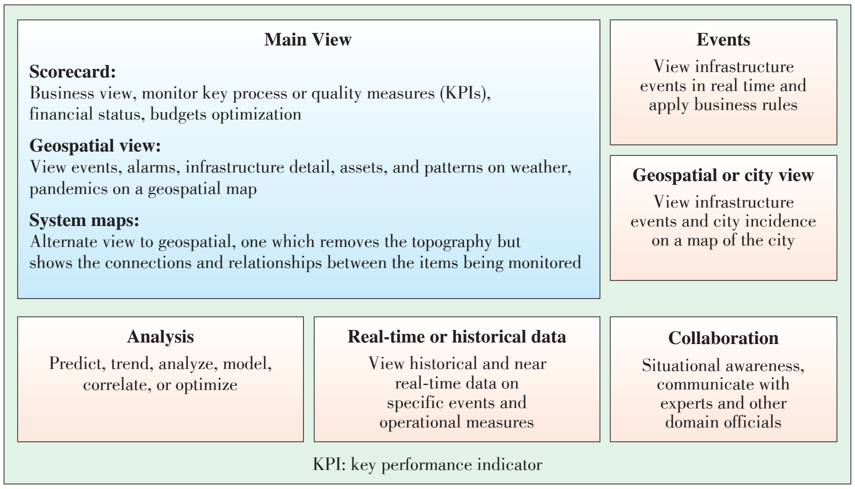
IBM IOC unifies applications and processes that are traditionally independent and isolated. It leverages real-time visibility of cross-city data to reduce cost; it anticipates and proactively manages problems to mitigate impact to services and citizens, and coordinates cross-agency operations with business and citizen participation to drive economic prosperity and increase citizen involvement.

The conceptual containment layout (**Fig. 1**) has the following key components:

1) Main viewer: The main viewer displays several views of city information. For example, a user can switch between:
- A view showing a heat-map presenting an at-a-glance visualization of the city key performance indicator (KPI) status
- A view of the KPI scorecard with a detailed textual display showing actual versus planned metric data

The main viewer presents two types of graphical information:
- Geospatial information, which is displayed on a map where various layers can be turned on or off depending on the level of detail and information required. The geospatial view is customizable to show information for a specific KPI.
- System maps, which are a schematic view of the linkages between various attributes.

2) Event widgets: Event widgets are the eyes and ears of the IOC, capturing event data at the infrastructure level and making it available to an enterprise service bus (ESB). Business and process rules can be applied to further enhance decision-making. The effect of KPIs can be determined in near real time. As events occur, business rules can be applied to determine which KPIs will be affected. Geospatial and city view: The city view widget can provide many overlays on a geospatial map including weather patterns, streets, buildings, events, infrastructure, work order, and asset detail. The city view widget provides key information. For example a public safety official can view the current police officer assignments or a water operator can review pipe infrastructure and current work order requests for the day.
3) Collaboration: The ability to instantly communicate and share information within the IOC is helpful in breaking down barriers. Interest groups, including managers and SMEs, can be created, their status can be instantly determined, and real-time contact can be made for a quick update.
4) Real-time and historical data: A key role of the IOC is to provide all pertinent information within a well-defined context. If a KPI has an attribute associated with it that captures data, it can then be displayed through the charting widget. With filters, the data can be viewed over a period of time or against other data sources. This information can be used to show trends, comparisons, correlations, and more. If required, the user can access the raw data and export that data to another application for further analysis.
5) Analytics: Much value is gained by a city harnessing the advantages of analytics. The IOC gathers data from many domains or departments, allowing the city to make informed decisions. As explained in "event widgets", a weather forecast of heavy rainfall, along with other factors (such as street layout and gradient) and related data (such as maintenance requests), can be used in the analysis of the situation. Ana-



▲**Figure 1. Conceptual illustration of IBM Intelligent Operation Center [19].**

lytic capabilities in the IOC provide information that a city manager can use to analyze the effect on services (such as street maintenance and bus routes). This information can be used to determine action that needs to be taken to control the extent of the flood, street closures, and other assets located in the affected area.

6) Task management: The IOC can be profiled and customized to suit the user's preferences. A list of tasks that require action by the user can be displayed, allowing the user to concentrate on what is required and increase their efficiency. The user can update and respond to pending tasks, such as performing impact analysis on a predicted scenario or responding to a directive that requires the attention of their department.

4.1 Mechanism: Event-Based Management

Event-based management contrasts with traditional authority discipline based management. In the case that a truck scattered some chemical on the road by accident, the bureaus of road, traffic, public safety and environment protection are required to take action immediately. In practice, coordination between those bureaus is not as smooth and efficient as we expect.

The following is an example of event-based management with intelligent operation center.

Scenario: Coordination of Resources in Response to Events on a College Campus

On the day of a home football game with a sellout crowd, a pipe bursts under a nearby dormitory. The university president or management would see:

- Visualization of water pipe event location on map, and in proximity to the football stadium, nearby dorms, and a bus route.
- Status of campus departments, including several that have an issue in need of attention (civil affairs, mass transit, water, and housing).
- Nearby video camera that shows a live view of area affected by the burst pipe event.

Anticipate Problems

1) Overall management problem identification and analysis

The University Manager:

- Views the water pipe event along with other nearby events/issues that may be effected
- Notices system notification that the event is in the same area as the football game and corresponding bus routes, which may require some intervention.
- Reviews current weather forecast to see if rain might affect water cleanup/repair and/or if it might further complicate crowd flow around the spill.
- Escalates the event to an incident to initiate a predefined cross-agency process for redirecting the crowd, rerouting the bus, and evacuating the dorm.

2) Identifying coordinate resources needed and making necessary arrangements

Campus Manager:

- Views the details of the water pipe event, and the related events at the dorm, stadium, and bus route.
- Reviews the status of the in-process workflow and is able to monitor progress as the appropriate police resources are redirecting the crowd, after the dorm has been evacuated and the bus has been rerouted.
- Collaborates with the water department to ensure that the water pipe is being repaired.

4.1.1 How Event Information is Delivered

Events are characterized as short, self-contained IT messages that have embedded information used by receivers to understand or become aware of the incident. Events can be published to topic queues and read by all interested subscribing IT systems. Event data comes from operational control systems and triggers processing in the IOC. One key part of the IOC is the event processing subsystem (EPS), which stores and manages the events.

4.1.2 Unified Event Data Format

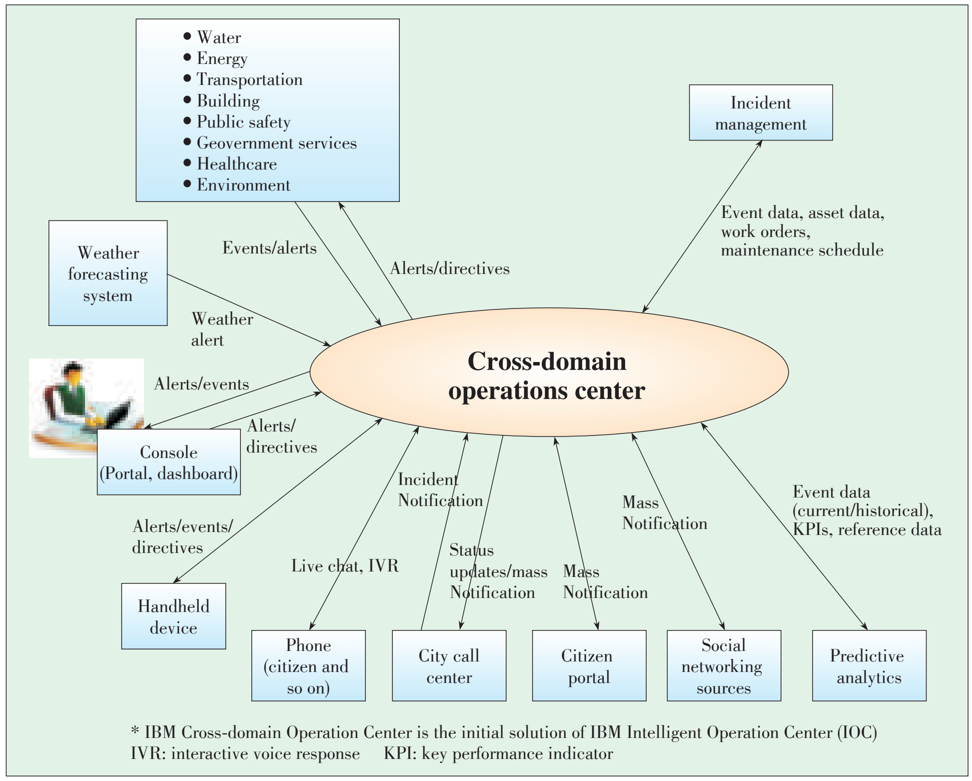
Event data comes from operational systems in various formats. The event data is normalized as it is received into a common format. The event normalizing protocol must be structured to fit the events found in city management systems and the IOC. A standard model manager can be used to provide a common dictionary of city assets and to map asset relationships. This tool enables effective analysis and response to events without the need for multiple translations of information.

4.1.3 How Event Alerts are Displayed

Events are self-contained IT messages that are maintained within the IOC. Alerts encapsulate one or more events and might require specific attention. For example, alerts affect the operation of city systems, such as water treatment or public transportation, and can report ad-hoc occurrences that affect city life. Alerts that affect the operation of a city are forwarded to the IOC from operational systems. At the IOC, the alert is displayed on an operations dashboard and can be analyzed to determine appropriate actions.

In some cases, the IOC can issue advice to the domain or department to consider the information being provided. Often an impact analysis statement is requested from various city domains to assess the overall impact on the city.

Another form of communication is notifications that are based on a subscribe/publish service, where citizens can be notified of information of interest or public alerts by using an SMS.



▲**Figure 2. Event information flow among systems [19].**

4.1.4 KPIs Based on Event Management

Knowledge of government policies of a city, including organizational and authorization structure, is important for determining IOC requirements. These policies must be understood at the domain and cross-domain level, enabling communication and coordination flows to be defined and implemented. In addition, the measurement of governance effectiveness is important to many cities and can help define and measure KPIs. Another important element is control mechanisms, which ensure compliance with various policies.

The tracking of events, alerts, directives, notifications, and advisories is required to manage incidents and situations. These city actions and communications (regardless of the form they take) can be captured in the form of an audit trail, for governance, and to enable learning

4.2 Data Collection and Access

4.2.1 Data Collection

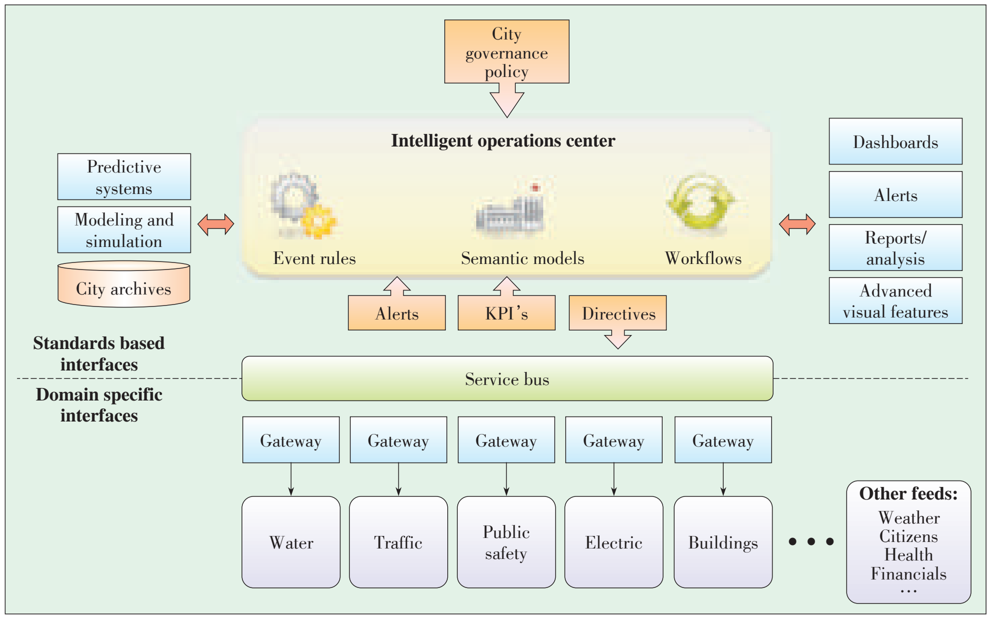
A cross-domain operations center provides a holistic view of the city by allowing access to information and data collected from a shared information space (**Fig. 2**). This shared information space contains information from various sources in the city and enables domains to contribute relevant data and analysis. This approach ensures that all related information is provided to city officials, giving them a comprehensive view of problems. It also enables them to understand and take action in a coordinated manner across city domains. The IOC receives data in a format that can be processed and updated by the system.

4.2.2 Data Access: Comprehensive and Role Tailored Dashboard

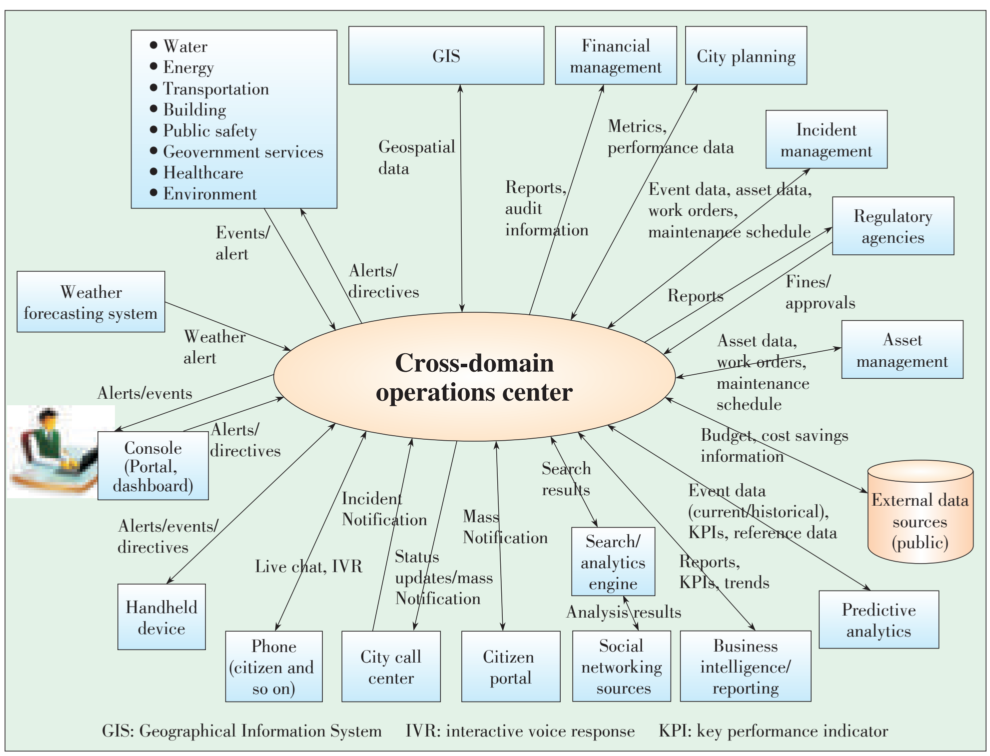
From a user-experience perspective, the central challenge is that cities are enormously complex and dynamic. The IOC must support users ranging from city executives to domain experts who have a deep expertise in particular aspects of a city's functioning. Furthermore, there should be an executive dashboard to depict the overall status of the city's operations. This spans individual agency-specific solution areas and enables drill-down capability into each underlying agency (e.g., water management, public safety and traffic management), and it provides for integrated collaboration within the views.

IBM IOC has a comprehensive web-based environment that provides consumable information that is necessary for making informed decisions. The dashboard provides a customized user interface based on specific user roles within a city organization or domain, such as the following views:

- City mayor or other top officials within the city. This view provides a high-level roll-up of key domain KPIs. By using this view, officials can collaborate on decisions and determine actions or directives for the city domains.
- City operator. This view is tailored to individuals or teams managing cross-domain events or incidents. These events or incidents can have a broader effect on the city as a whole.



▲Figure 3. Overview architecture of IBM IOC [19].



▲Figure 4. System interactions of IBM IOC [19].

- Incident or emergency manager. This view helps managers understand an incident or emergency, manage the response, and track it to completion.
- City domain manager or operator. This view helps managers or operators track and manage KPIs, events, and work orders. They can view current and historical data and participate in cross-organization collaboration.
- Citizen. This view can be informative, e.g., provide details about road closures, and encourage citizen input, such as real-time information about crime and city infrastructure issues. This interaction is intended to be motivational and to encourage involvement in sustainable programs, such as reducing water or electricity use.

4.3 Comprehensive System Infrastructure

IBM IOC as a system of systems supports:

- Integration of subsystems: Standard interfaces and domain-specific interfaces for integrating various systems (**Fig. 3**), integration of various reporting/alerting/analyzing mechanisms, and portal and user account management system to ensure unified management and scheduling across various systems.
- Application subsystems: (**Fig. 4**) Integrate subsystems such as transport, environmental protection, government information library, etc.

Relying on the IOC, IBM also provides systems for vertical applications of intelligent management of all aspects of the city, including government services, public security, intelligent transportation, water management, digital city management, food safety, and logistics.

In terms of cross-domain operation and management, IBM IOC supports: organization-wide dashboards, domain analytics, event and KPI management, geospatial mapping, data modeling and integration, simulation and visualization, cross-department collaboration, situational awareness, incident management, alerts, and directives.

4.4 Citizen Engagement

By coordinating across agencies and collaborating with citizens and new partners, cities can transform traditional work structures to promote innovation.

The IOC gets data from citizens by leveraging IBM citizen sensing platform. This platform works in three ways:

- It helps the city understand its citizens. A city planner can understand the needs and priorities based on direct citizen input, effectively deputizing thousands of people to report problems and ask questions. This saves staffing costs and improve citizen engagement.
- Timely updates to citizens and evidence of actions taken against their requests to engage citizens in the government process. The goal is more involved and more informed constituents that are satisfied with the services being provided.
- The IOC can interact with citizens and visitors in numerous ways. This communication and interaction can be bidirectional between the city and citizens. The city can provide citizens with important safety information and information related to city policies. For example, if a serious traffic accident is blocking a main road through the city, the city can notify the citizens. As another example, if a hazardous waste is spilled and people in a specific area need to be evacuated, the city can notify citizens in that area. In addition, citizens can report up-to-date situational information to the city. With this information from citizens, city officials can revise decisions or allocate resources as needed.

The IOC also provides citizens with a platform to report less serious incidents, such as water leaks or potholes in their area. This information can be used to schedule maintenance. Services can be provided to citizens so that they can subscribe to notifications pertaining to particular areas of the city. For example, public traffic notifications can identify unscheduled road works or other unplanned disruptions. These notifications might allow citizens to adjust their travel plans, maybe take a different route or means of transport to work.

The IOC can also be used to develop programs and initiatives with citizens. For example, if water or electrical meters share usage data with the IOC, this data can be used to build usage patterns for individuals, streets, and districts. Motivational techniques can be used to help manage this usage. These techniques can include pricing models or incentives to those who participate in the program and actively reduce resource consumption. This approach has a positive benefit for the city by reducing dependency on resources and for the citizens who are making a difference and potentially getting paid for it.

4.4.1 A Practical Case of Citizen Collaboration: City Forward

The City Forward [20] is a free, web-based platform for analyzing and visualizing data from cities around the world. Right now, the City Forward offers municipal data for more than 100 world cities, and it is growing.

That data is publicly available, but it is often scattered or exists in a variety of formats, making it hard to compare one city or service to another. Even in a single city, such data is often published independently by individual agencies, making it hard to see the bigger picture. The City Forward addresses these issues by bringing useful statistics and graphing tools together in one place, offering easier and more insightful analysis. Citizens can easily access to those data and custom their own analysis.

5 Intelligent Operation Center Case Study: Rio de Janiero

Rio Emergency Management Centre, Brazil: The IBM IOC powered Emergency Operations Center provides real time alerts for floods and landslides to prevent loss of life and infrastructure, better inter-agency coordination, better management of resources and better preparation for addressing emergency and disaster management situations. Rio de Janiero City Hall deployed a Web-based portal that integrates information, applications and resources across agencies. It reduces fully deployed disaster response times from days to hours. [21], [22]

- 2009: Rio de Janeiro chosen for the 2014 World Cup and the 2016 Olympics
- April 2010: Floods result in serious loss of life and property
- May 2010: In consultation with IBM, visionary Mayor decides to not only address flood crisis management, but to integrate city's operations across all departments

5.1 Pain Points

The city has many pain points before using the IBM IOC solution, for examples:

- aging systems that were in silos
- no common operational picture across disaster and crisis management teams
- legacy alert system that was very manual intensive
- lack of comprehensive disaster management plans
- difficulties in organizing and distributing disaster management plans
- Decision making process for flood prediction was based on basic weather reports and radar information, which was not sufficient to have good predictions.
- lack of visibility of resources and overall picture of resource requirements with wide spread disaster
- lack of radio interoperability across departments

5.2 Solution and Implementation

The solution and implementation include the following components or aspects:

- common operation picture with data fusion in City Operations Center with Geographical Information System (GIS)
- dynamic planning tools
- consultancy to define and deliver disaster management plans
- load Operations Center planning module with new disaster plans
- partially automate legacy alert notification system within City Operations Center
- cross agency collaboration with collaboration suite

- incident management tool
- recovery management tools
- integrate legacy video feeds
- future phase considerations—compatible with City Operations Center based on IBM's Government Industry Framework
- radio interoperability with Radio Connect
- full alert notification automation and GIS based alerting
- dynamic team building to do intelligent assembly of first responder teams from available personnel
- process automation and Re-engineering
- include smarter transportation elements
- flood and water level sensors
- smart Video analytics.

5.3 Accomplishment

Inaugurated in December 2010, the 400 member strong operation center consolidates 30 municipal agencies, ranging from emergency services to sanitation and public transport, into a single, integrated command center that can monitor weather, traffic, and aspects of the city and react swiftly and effectively in face of emergencies and crises and would do so through cutting edge technology. These cutting-edge technologies have allowed the Rio de Janeiro to not only address problems it faces from severe storms but also addressed other lagging areas of the city's public infrastructure [22]. The benefits include:

- Using the Operations Center's website, Brazilians can get real time weather updates and traffic advisories (both for public transportation and for the roads in general) for the whole city.
- In addition, all Brazilians can view live feeds via the Operations Center's web page for many of the streets in Rio de Janeiro. On the ground, storm warning systems much like Tsunami warning systems alert people to severe storms for their neighborhoods and emergency tests/simulations.
- Going beyond even having a standard web presence, The Operations Center maintains both Facebook and Twitter accounts that regularly send out alerts and updates through the day and night.
- As a final and important aspect of the Operation Center's public accessibility is its ability to have its alerts available to Brazilians on the move during their day.
- Considering that smartphones which can easily access the Operations Center and its updates via the website and social media apps like Facebook and Twitter only account for under 10% of the current mobile phone market for Brazil, nearly all Brazilians in the larger cities have some kind of cellular phone, all the same updates that are available for non-smartphones and emergency messages are sent out to all citizens in the affected area through SMS messages, making it so that even a person still using a monochromatic display cell phone (think the greenish-grey phone displays) can still receive vital messages such as incoming storms, evacuation orders, and evacuation routes on their mobile phones.

6 Conclusions

The city, as the container of human civilization, is expanding to its limits. While the shape of this container was decided centuries ago, reflecting the knowledge and desire of our forefathers. It is our chance and duty to re-define the shape, with technology tools and deeper understanding of the society we live in. That shape of container increases social efficiency, convenience and happiness of citizens.

In addressing the urban operational collaboration problems, efforts are being made on capitalizing on new insights by distilling insights from the massive quantities and sources of data generated by day-to-day operations, cities can approach longstanding challenges in new ways; creating system-wide efficiencies by optimizing and integrating operations, cities can leverage interdependency between systems to do more with less. Collaborating in new ways by coordinating across agencies and collaborating with citizens and new partners, cities can transform traditional work structures to promote innovation.

Looking into the future of city operation, we expect more flexible, interactive urban operation organizations instead of separated bureaus. We will have unified systems and well-managed data instead of information islands, sharper sense to danger and accident, more accurate predictions, more effective and timely actions, and finally, more democratic cities that listen to people, respect their demands, and protect their rights.

Acknowledgment

The authors would like to thank IBM Smart City Community for their excellent work which forms the foundation of this paper. We would also like to thank APEC Smart City Forum Organization Committee to invite one of us to present IBM Smart City architecture and solutions in both 2012 and 2014 APEC Smart City Forums, held in Changzhou, China [23].

References

[1] IBM US Official Website. (2015, Jul. 20). *Smart Planet* [Online]. Available: http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/smarterplanet/
[2] L. Munford, *The Culture of Cities*, 1st simplified Chinese edition, translated by J. Song, X. Li, and M. Zhou. Beijing, China: China Architecture & Building Press, 2009, pp. xvii.
[3] Xu Yuan. (2015, Mar. 31). *Cheng Shi de Qi Yuan III: Zhu Cheng Yi Wei Jun, Zao Guo Yi Ju Min* [Online]. Available: http://www.thepaper.cn/newsDetail_forward_1316329
[4] W. Fan and M. Cai, *Chinese History*, Beijing, China: People's Press, 2009.
[5] L. Mumford, *The City In History: A Powerfully Incisive and Influential Look at the Development of the Urban Form through the Ages*, Simplified Chinese edition, translated by J. Song and W. Ni. Beijing, China: China Architecture & Building Press, pp. 69.
[6] Xi'an Cheng Qiang. (2015, Jul. 20) [Online]. Available: https://zh.wikipedia.org/wiki/西安城墙

[7] N. Spencer and D. Butler. (2010, Oct. 20). Cities: The century of the city. *Nature* [online]. 467, pp. 900 – 901. Available: http://www.nature.com/news/2010/101020/full/467900a.html?message=remove&s=news_rss
[8] London. (2015, Jul. 20) [Online]. Available: https://en.wikipedia.org/wiki/london
[9] 5th Ring Road (Beijing). (2015, Jul. 20) [Online]. Available: https://en.wikipedia.org/wiki/5th_Ring_Road_(Beijing)
[10] L. Mumford, *The City In History: A Powerfully Incisive and Influential Look at the Development of the Urban Form through the Ages*, Simplified Chinese edition, translated by J. Song and W. Ni. Beijing, China: China Architecture & Building Press, pp. 31–33.
[11] Baidu Baike. (2015, Jul. 20). *Shi Ye Dan Wei* [online] Available: http://baike.baidu.com/view/65492.htm
[12] L. Mumford, *The City In History: A Powerfully Incisive and Influential Look at the Development of the Urban Form through the Ages*, Simplified Chinese edition, translated by J. Song and W. Ni. Beijing, China: China Architecture & Building Press, 2014, p. 563.
[13] M. M. Grynbaum. (2010, Mar. 23). *Gridlock May Not Be Constant, but Slow Going Is Here to Stay* [Online]. Available: www.nytimes.com/2010/03/24/nyregion/24traffic.html?r=0
[14] China Internet Information Center. (2015, Jun. 5). *China's 1st MERS Patient Remains in Serious Condition* [Online]. Available: www.china.org.cn/china.2015-06/05/content_35744374.htm
[15] Wikipedia. (2015, Jul. 28). *Boston Marathon bombing* [Online]. Available: https://en.wikipedia.org/wiki/boston_Marathon_bombing
[16] A. Pentland, *Social Physics: How Good Ideas Spread—The Lessons from a New Science*, Simplified Chinese ed, translated by X. Wang and R. Wang. Zhejiang, China: Zhejiang People's Publishing House, 2015, p. 136.
[17] Wikipedia. (2015, Mar. 1). *DRAPA Network Challenge* [Online]. Available: https://en.wikipedia.org/wiki/DRAPA_Network_Challenge
[18] E. Stepanova. (2011, May). *The Role of Information Communication Technologies in the "Arab Spring"—Implications Beyond the Region* [Online]. Available: www.ponarseurasia.com/sites/default/files/policy-memos-pdf/pepm_159.pdf
[19] Redbook. *Smarter Cities Series: Introducing the IBM City Operation and Management Solution* [Online] Available: www.redbooks.ib,.com/redpapers/pdfs/redp4734.pdf
[20] Chicago Architecture Foundation, *Big Data* [Online]. Available: www.bigdata.architechture.org/highlights
[21] IBM, "Smart city case study," Internal Report.
[22] *RIO de Janerio and IBM's Smarter Cities Project* [Online]. Available: www.museumofthecity.org/rio-de-janerio-and-ibms-smarter-cities-project/
[23] Weidong Kou, "IBM's smart city architecture and solutions," *Keynote Speech at APEC Smart City Forum*, Changzhou, China, Apr. 2012.

Manuscript received: 2015-08-16

Biographies

Rui Cao (caoruibj@cn.ibm.com) currently is IBM Global Business Service consultant. She received a Bachelor degree in industrial design from Beijing University of Posts and Telecommunications, and a Master degree in engineering from The University of Hong Kong, with major in E-commerce & Internet Computing. In the past three years, she had been with IBM, working as a consultant in smart city, banking, telecommunication, and retail industries, with emphasis on applying advanced information technology into traditional industries.

Weidong Kou (kouwd@cn.ibm.com) currently is IBM Asia Pacific & Great China Group Chief Architect and Cloud Executive for Global Alliance. He has 30 years of IT industrial and academic experience, including over 20 years of IT practice in IBM and over 6 years of application software development in AT&T and Siemens. He was the dean of the school of computer sciences and engineering at Xidian University, and he has been a full visiting/adjunct professor in 17 universities around the global, including the University of Maryland in US and the University of Hong Kong. He has published 7 books and over 100 papers in journals and international conferences. Dr. Kou is the Laurent of the highest award from China National Government to the international experts in 2004.

DOI: 10.3969/j. issn. 1673-5188. 2015. 04. 004
http://www.cnki.net/kcms/detail/34.1294.TN.20151130.0926.002.html, published online November 30, 2015

A Novel Data Schema Integration Framework for the Human-Centric Services in Smart City

Ding Xia, Da Cui, Jiangtao Wang, and Yasha Wang
(Peking University, Beijing 100871, China)

Abstract

Human-centric service is an important domain in smart city and includes rich applications that help residents with shopping, dining, transportation, entertainment, and other daily activities. These applications have generated a massive amount of hierarchical data with different schemas. In order to manage and analyze the city-wide and cross-application data in a unified way, data schema integration is necessary. However, data from human-centric services has some distinct characteristics, such as lack of support for semantic matching, large number of schemas, and incompleteness of schema element labels. These make the schema integration difficult using existing approaches. We propose a novel framework for the data schema integration of the human-centric services in smart city. The framework uses both schema metadata and instance data to do schema matching, and introduces human intervention based on a similarity entropy criteria to balance precision and efficiency. Moreover, the framework works in an incremental manner to reduce computation workload. We conduct an experiment with real-world dataset collected from multiple estate sale application systems. The results show that our approach can produce high-quality mediated schema with relatively less human interventions compared to the baseline method.

Keywords

schema matching; schema integration; smart city; human-centric service

1 Introduction

Human-centric service is an important domain of the smart city and includes rich applications that help residents with shopping, dining, transportation, entertainment, and other daily activities. While offering services, these systems have generated a large amount of hierarchical data. Usually data from each system is incomplete, and one system complements another. Take the data of second-hand housing for example. There are many second-hand housing information sharing application systems containing data for a given city in China. In Beijing, such systems include lianjia.com, 5i5j.com, 58.com, fang.com, iwjw.com, and many other local forums. Since each system only contains certain information, a resident who wants to buy a second-hand house or apartment needs to browse the systems one by one and pull together all the parts of the information by themselves. In another scenario, if a city planning or market investigation department wants to know the situation of the second-hand house market of the whole city, they also need to integrate data from different systems. As the data schemas of different systems are diverse, the schema integration, whose goal is to establish a whole and unified schema for all the multi-sourced datasets, is necessary and crucial.

Traditional schema integration techniques are usually used in scenarios where the number of data schemas is small, the structure or semantic of the schemas is well understood, and the linguistic resources for semantic matching across different schemas are sufficient. Typical application scenarios include the evolution of product directories in the domain of e-commerce, data system integration caused by the company merges, and so on [1]–[5]. However, data integrating in the domain of human-centric service of smart city is much more challenging due to its distinct characteristics as follows:

1) Broad application domains and lack of domain knowledge. Human-centric service is about almost everything in a resident's daily life. There has not been any standard or knowledge base providing support for the semantic matching in this domain. For example, in second-hand housing data sets, there are dozens of terms for the sales agent, such as agent, broker, advisor, secretory, or housekeeper across various systems. Therefore, we need a domain specified term dictionary to help us figure out whether two elements from two schemas

This work is funded by the National High Technology Research and Development Program of China (863) under Grant No. 2013AA01A605.

refer to the same concept. However, no such dictionary is available, and it is time-consuming to build one. Also, traditional string-based matching algorithms perform poorly for Chinese labeled elements.

2) Large number of schemas. Even in a fine-grained sub domain such as second-hand housing, there are dozens of systems from which data schemas need to be integrated. Traditional schema integration methods mostly involve studying how to match or integrate two schemas, and if we simply use them to work on every pair of many schemas, it will cost too much time and has poor extendibility.

3) Label Incompleteness. On the one hand, some data in human-centric service domain is acquired from web tables and sometimes the labels of schema elements are missing. This makes traditional element-level matching techniques unsuitable. On the other hand, instance data is usually available and can be used to assist schema matching. However, instance data from different systems do not always overlapped, so we won't get satisfying result if we simply calculate instance overlapping level to represent similarity between two elements.

Due to the abovementioned characteristics, the existing approach is incapable of handling the data schema integration for human-centric service in smart city. Therefore, we propose a novel approach of schema integration with data from domain of human-centric services in this paper. In our approach, we use a mediated schema to help integrate multiple schemas in a quick manner. Every schema will be matched and integrated to the mediated schema only once, i.e., one iteration, and the mediated schema is updated and extended after each iteration. During each iteration, a depth-first-search algorithm is used to control element-matching and integration order. Five matchers that utilize both schema metadata and instance data are combined for schema matching. We introduce a similarity entropy based interactive method of human intervention controlling to make matching results more precise. After schema matching, a set of conflict resolution strategy is adopted to solve all kinds of complex conflicts and then form a better and more complete new mediated schema.

The rest of the paper is structured as follows. Section 2 reviews related work. Section 3 gives the framework of our approach. Section 3 describes detailed algorithms. Section 4 designs and conducts experiments to do validation. Section 5 concludes this paper and analyzes future work.

2 Related Work

Schema integration can be divided into two parts: schema matching and mediated schema generation. Schema matching involves forming mapping between elements in different schemas, which have the same or similar semantics, while the mediated schema generation is to generate a whole and unified schema for all the integrated schemas based on the result of schema matching.

2.1 Schema Matching

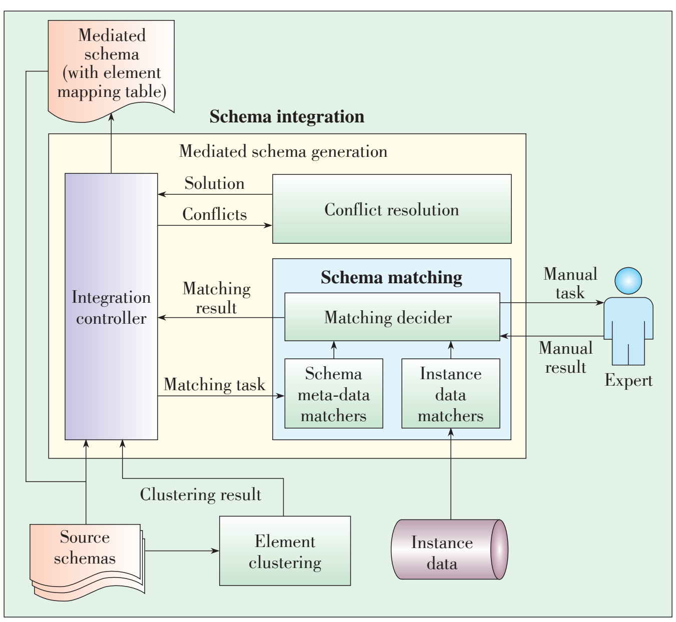
According to the input, schema matching can be divided into two categories: metadata based matching and instance-based matching. Metadata-based matching uses the metadata, including labels and structures of the elements in the input schemas, as the input of the matching. The most basic matching algorithms in this category are called element-level matchers, which maps the elements from different schemas based on the their labels according to the string similarity [1], [2], [6], linguistics-based semantic relationship [7], [8], or word co-occurrence in schemas [9], [10]. Another kind of metadata-based matching algorithm is called structure-level matchers, which not only take the element information into consideration but also the structure of the element information. Typical structure-level matchers include graph-based matchers [11], [12] and path-based matchers [3], [13]. Although metadata-based matching algorithms are fast, they may be unfeasible when the metadata of the schemas are incomplete. On the other hand, instance-based matching algorithms do not depend on the metadata of schemas. Instance-based matchers dig similarity between elements from instance data. Typical instance-based matchers calculate the similarity among elements according to instance statistical features [14] or the overlapping instances [15], [16].

2.2 Mediated Schema Generation

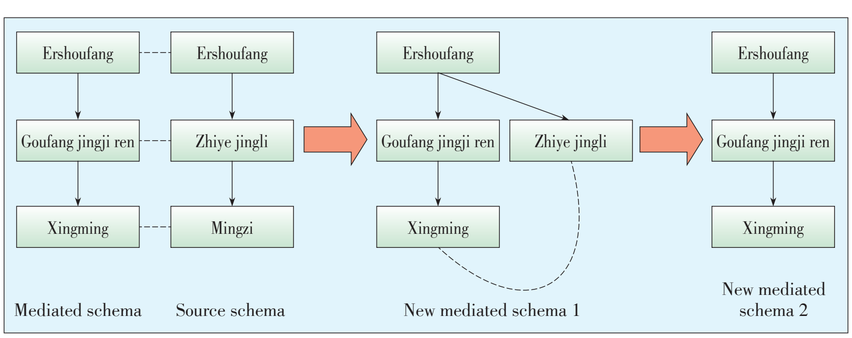
Based on the results gain from the schema matching, mediated schema generation algorithms try to resolve conflicts in element naming, definition, and structure inconsistency from different schemas. Then, such algorithms form a mediated schema containing all of the elements in the integrated schemas to make the heterogeneity of schemas transparent to the data users [4]. In addition to the conflict resolution, the number of input source schemas has great impact on mediated schema generation. Traditional resolution, such as XSIQ [4], is applied to two source schemas. However, when there are many schemas to be integrated, techniques that apply to two schemas greatly increase complexity. XINTOR [17] provided a global schema integration technique for multiple source schemas. It uses the statistical characteristics of element structure to generate a mediated schema. XINTOR greatly increases the efficiency of matching multiple schemas but it has poor extensibility. Another idea is taking mediated schema as an intermediate product and integrating all the other schemas into it. This increases both efficiency and extensibility. In some research [18], human experts are required to develop a mediated schema. However, under the background of human-centric services which have massive multi-source heterogeneous datasets, these techniques pose too much burden on experts. Automatic generation of mediated schema will avoid large number of human effort, for example, PORSCHE [19]. PORSCHE integrates one schema with mediated schema for one time and in the meantime updates the mediated schema. It only takes advantage of schema information, which results in limited accuracy. Moreover, it has simple conflict resolution strategy and some conflict types are not considered. Hence, the quality of mediated schema generated by PORSCHE is low.

3 Framework Overview

The framework of our approach is shown in **Fig. 1**. There are three kinds of input into the system. The first kind is schema metadata of the data sources, which defines the structure of the data, and the elements making up the structure. For example, the source schema in **Fig. 2** is part of the schema metadata of the web application called 'Anjuke', and it consists of three string labels. The second kind is the instance data of every leaf element, providing the actual useful information, such as the leaf element 'Mingzi' in the source schema in Fig. 2. The instance data are real names of second-hand housing agents. The third kind is human intervention. The major output is the mediated schema. In our approach, the mediated schema serves as



▲Figure 1. Approach framework.



▲Figure 2. Nested path conflict type three.

intermediate product and all the other source schemas will be matched and integrated to it sequentially, resulting in its update and expansion.

Element clustering module takes all the source schemas as input and uses element level matcher to calculate the similarity of any two elements from different schemas. Based on these similarity values, all elements will then be classified into several clusters, which are used to reduce the complexity of the following computation.

The schema integration module takes source schemas, current mediated schema, and instance data as input to complete schema integration task under the support of experts' knowledge and elements clustering result. In this module, first, one of n source schemas is chosen as the initial mediated schema. Then, the left n-1 source schemas are integrated into the mediated schema sequentially in an n-1 iteration. Each iteration is a complete schema integration process: the integration controller submodule traverses the matching source schema in depth-first-search order and for every element that is traversed and finds its candidate matching elements in the mediated schema using the clustering result. Then schema matching submodule calculates similarities between the traversed element and its candidates only, thus avoiding the similarity calculation of all element pairs. In the schema matching submodule, element level matcher, ancestor path matcher, tree edit distance matcher, which make use of schema metadata, and statistic based instance matcher, content based instance matcher, which make use of instance data are used. Some of these matchers are modified or redesigned to adapt to Chinese labels, and different matchers are reasonably and efficiently combined according to their characteristics. Also, human intervention based on similarity entropy is introduced. Questions are generated and sent to experts if the schema matching algorithm cannot automatically decide which candidate is the most appropriate one to the traversed element. The matching decider submodule sends these questions, gathers experts' feedback, and determines the final matched element pair. When matched, the conflict resolution submodule will be used to solve conflicts between the two matched elements. The mediated schema is updated and expanded after every iteration, and at the end of the last iteration, the final mediated schema and element mapping table is obtained.

Only an element clustering process and another iteration process of schema integration are needed when a new source schema is added in. All the updates and expansions brought by the new added source schema are incremental and the pre-existing results of schema integration will not be reversed.

4 Algorithm Design

Here, we describe the algorithms of the four most important modules: element clustering module, integration controller submodule, schema matching submodule, and conflict-resolution submodule respectively. In schema matching submodule, the five different matchers, the combination design and the human intervention design are introduced in order.

4.1 Element Clustering

This module clusters elements according to the similarity calculated by the element level matcher. We use and modify the Kruskal algorithm which is originally used to calculate the minimum spanning tree, to do the clustering job. First, we calculate the similarity values of every element pair using the element level matcher, which is the quickest among all matchers. Then we sort them in ascending order and every time pick an element pair that has the smallest similarity to be clustered. The intuition behind this algorithm is to make the difference among clusters to be as large as possible.

4.2 Integration Controller

During each round of the iteration, the current source schema is matched and integrated into the mediated schema. We traverse the current source schema in depth-first-search order to ensure that whenever an element is about to be matched, its father element is already matched. The detailed steps are described as follows:

1) The root elements of all source schemas should all be matched because all the source schemas are from the same domain. We traverse the current source schema from the root's first child element.
2) Denote the currently traversed element as element a. Find the cluster containing a from the clustering result, denoted as A. If the size of A is one, go to step 7.
3) Extract all elements which exist in the mediated schema from A to form another set B. If the size of B is zero, go to step 7.
4) For each element bi in B, invoke schema matching submodule to calculate its comprehensive similarity with a, denoted as Si, i is from one to the size of B.
5) Calculate $Entropy(a)$ using $S1$ to $S|B|$. If $Entropy(a)$ is greater than $T*ln|B|$, send question to experts to receive human intervention and get the right matching element bm, otherwise, choose the element that has the greatest comprehensive similarity with a, denoted as bm as well.
6) Match a to bm and invoke conflict resolution submodule to solve possible conflicts. Update the mediated schema and record elements mapping relationships. Go to step 8.
7) If no element in the mediated schema can be matched to a, find the father element of a, denoted as fa. Denote the element that has been matched to fa as fb, and add a to the mediated schema as fb's right-most child element.
8) If a is not a leaf element, get the first child element of a and repeat step 2. Otherwise get a's next brother element and repeat step 2. If a is the last element in depth-first-search order, algorithm ends.

4.3 Schema Matching

In our approach, we use five matchers for schema matching. Both schema metadata and instance data are utilized, and different matchers are reasonably and efficiently combined according to their characteristics. Interactive human intervention based on similarity entropy is introduced to improve matching accuracy.

4.3.1 Element Level Matcher

Traditional element level matchers primarily acquire similarity value between elements according to their string-based similarity and semantic-based similarity. Commonly used string-based algorithms perform poorly whereas processing Chinese label, and semantic-based algorithms rely deeply on external domain synonyms dictionary. Therefore, we use Word2Vec [20] to help calculate element level similarity. Word2Vec is a tool to map words to K-dimension real vector space, so the similarity between two words can be represented by the cosine value of the two K-dimension vector. If trained by a large corpus in a particular domain, similarity values calculated can not only represent the string-based similarity of words but also cover parts of their semantic-based similarity. This matcher runs fast and can be served as a standalone matcher in most cases, but it only has an ordinary accuracy, not satisfying enough. **Algorithm 1** shows the process of this matcher.

Algorithm 1. Element level matcher

Input: : $s1$, $s2$: labels of two elements to match
Output: *Element_Similarity* : element level similarity

```
begin
  TokenList1 ← Tokenize(s1)
  TokenList2 ← Tokenize(s2)
  Element_Similarity ← 0
  for each token1 ∈ TokenList1 do
    for each token2 ∈ TokenList2 do
      Element_Similarity += Word2VectorSimilarity(token1, token2)
    end for
  end for
  Element_Similarity /= sizeof(TokenList1)*sizeof(TokenList2)
end
```

4.3.2 Ancestor Path Matcher

Ancestor path matcher is a kind of structure level matcher. The idea of this matcher is that two elements are similar to each other if they have matched ancestors and the paths from the matched ancestor to them are short enough. This matcher strictly control the matching relation of elements based on their ancestor elements' matching relation. By using this matcher we can effectively avoid mismatching case such as falsely matching an element labeled 'Xingming' whose father element is labeled 'Zhiye Jingli' to another element labeled 'Xingming' whose father element is labeled 'Yezhu'. **Algorithm 2** shows the process of this matcher.

Algorithm 2. Ancestor path matcher

Input: : a , b : two matching elements
Output: *AncestorPath_Similarity* : ancestor path similarity

```
begin
  c ← parentOf(a)
  AncestorPath_Similarity ← 0
  while c ≠ null do
    d ← MatchingNodeOf(c)
    if d.isAncestorOf(b) = true
      then break
    else
      c ← parentOf(c)
    end if
  end while
  AncestorPath_Similarity ← 1/length(c, a)
end
```

4.3.3 Tree Edit Distance Matcher

Tree edit distance matcher is another kind of structure level matcher. The idea of this matcher is that two elements are similar if their subtrees are alike. To measure the likeness of two trees, we use tree edit distance, the definition of which is the minimum number of operations required to transform one tree into another [12]. We use dynamic programming to calculate it. This matcher runs slowly but can measure the similarity between elements according to their subtree structure. For example, elements labeled 'Zhiye Jingli' and 'Goufang Jingjiren' can be very hard to match using other matchers, such as element level matcher. However, we can obtain a pretty high similarity through tree edit distance because the two element both have child elements labeled 'Xingming', 'Dianhua', 'Gongsi'. One obvious weakness of this matcher is that useless for leaf elements, which have no child element.

4.3.4 Statistical Based Instance Matcher

A statistical-based matcher is a kind of matcher that uses element instance data and focuses on statistical indexes of this data. We first calculate Eigenvectors of every leaf element using its instance data, then we use the Back Propagation Neural Network Algorithm to get a classifier for every source schema each.

Thirteen features are chosen for Eigenvector. For instances of numeric type, features are computed using its value, while for instances of entity type, features are computed using its length. We have six data type features: integer, floating number, URI, date, string and text, and seven statistical feature: maximum, minimum, mean, standard deviation, mean-squared difference coefficient, number of bytes and precision.

When calculating the similarity value between two leaf elements a and b, we will put the eigenvector of a and b into the classifier of schema A and B respectively to get two similarity values, then we take the average of them as the statistical based instance similarity between a and b.

This matcher uses element instance data, but for instances of entity type, it only takes their lengths into account and neglects their content. This matcher has high recall rate but low precision rate when used alone, and it runs fast.

4.3.5 Content Based Instance Matcher

Content based matcher is another kind of matcher that utilizes elements' instance data, it focuses on the content of elements' instance data. For two leaf elements, the overlapping level between the contents of their instance data can represent a sort of similarity between them. We first classify elements into three data types: text, entity and number. Then we calculate the content based instance similarity between two leaf element a and b as follows: 1) If a and b differ in their data type, the similarity value is 0. 2) If their data types are both text, we merge their instance data into two documents, count word frequency separately, and use the vector space model to convert the two documents to two vectors to calculate cosine value as their similarity value. 3) If their data types are both entity, for every instance of a, find the instance of b that has the smallest edit distance with it. We accumulate these smallest edit distance values and average it to get the similarity value between a and b. 4) if their data types are both number, we calculate their standard deviations and means of data instance separately, and acquire the similarity value between a and b as follows:

$$Content\_Similarity = 2/\left(\frac{|std1-std2|}{std1+std2}+\frac{|avg1-avg2|}{avg1+avg2}\right) \quad (1)$$

This matcher uses element instance data and focuses on the content of instance data. Due to the impact of instance subset problem, this matcher is precise but has a low recall rate when used alone. It runs slowly because a lot of computation is needed when getting the overlapping level.

4.3.6 Matcher Combination Algorithm

So far, we have introduced five matchers with different characteristics. Element level matcher runs the fastest but is not accurate enough, ancestor path matcher cannot be used alone but is a very good supplement, tree edit distance matcher is useless for leaf elements, and statistic based instance matcher and content based instance matcher both apply to leaf elements only. How to reasonably and efficiently combine them to get a comprehensive similarity for element pairs is a crucial problem concerned in this section.

Element level matcher has the highest calculating efficiency, so we use this to help with element clustering. The other matchers are less efficient and only work on parts of the whole element pairs based on the clustering result, thus the total calculation is greatly reduced.

The two matchers that use element instance data are complementary, so we can combine them in advance. Statistical based instance matcher has high recall rate and low precision, and a high similarity calculated by this matcher is actually not so reliable. On the contrary, content based matcher has high precision rate and low recall rate, and a low similarity calculated by this matcher is actually not so reliable. Considering that statistical based matcher is much quicker than content based matcher, we can first use statistical based matcher to get a similarity, and only when this similarity is high enough will we use content based matcher to adjust the unreliable result. This combination method can guarantee both a high accuracy and a high efficiency. We call this similarity instance similarity, and it can be calculated as follows, *Sta_Sim* is the statistical based matcher, *Con_Sim* is the content based matcher and *d* is an empirical threshold.

$$Instance\_Similarity(a,b) = \begin{cases} Sta\_Sim(a,b), & Sta\_Sim(a,b) < d \\ d + (Sta\_Sim(a,b) - d) * Con\_Sim(a,b), & Sta\_Sim(a,b) \geqslant d \end{cases} \quad (2)$$

We then combine instance similarity with the result of tree edit distance matcher because the former applies to leaf elements only whereas the latter is useless for leaf elements. When two matching elements are both leaf elements, the instance similarity between them is calculated; otherwise, we use tree edit distance matcher to determine the similarity. We denote this combined similarity as subtree similarity. As for the other two matchers—element level matcher and ancestor path matcher, we call their result as element level similarity and ancestor path similarity.

At last, we get the comprehensive similarity of every element pair by calculating the weighted mean value of subtree similarity, element level similarity and ancestor path similarity.

4.3.7 Human Intervention Design

The comprehensive similarity calculated above sometimes is still not accurate enough, so we introduce human intervention to our framework. We assume that experts' advice is always right but the number of expert queries must be controlled.

For every matching element in a source schema, there are several comprehensive similarity values that belong to it and its candidates. The idea of our human intervention is that if one of these similarity values is significantly higher than others, then it is the match. However, if these similarity values are so close, we cannot determine a prominent one. In the latter case, we can form questions and send them to experts to get a real match.

We use similarity entropy to decide whether there is a need to query experts. The equation to calculate the similarity entropy is as follows:

$$Entropy(x) = -\sum_{j=1}^{k} p(sim_j) * lnp(sim_j),$$
$$p(sim_j) = \frac{sim_j}{\sum_{i=1}^{k} sim_i} \quad (3)$$

For a candidate set consisting of K elements, *Entropy(x)* ranges from 0 to lnK. We import a threshold T and we ask experts for advice only if *Entropy(x)* is greater than $T*lnK$, otherwise we will choose the element that has the greatest comprehensive similarity with the matching element.

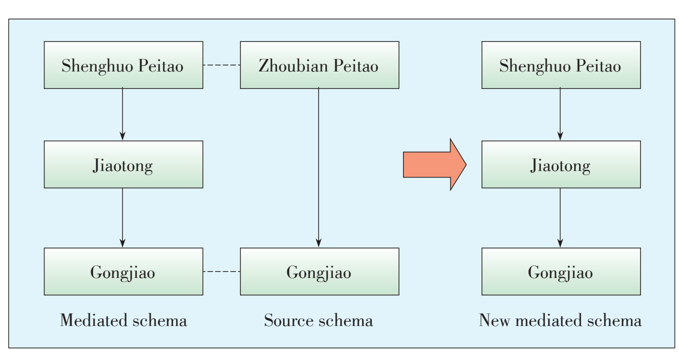
4.4 Conflict Resolution

Conflicts may arise when elements get matched. Conflicts and our solution strategies are as follows:

1) Synonym conflict. Our resolution is to reserve old element's label to be the one in the mediated schema but also record the new matching element's label for later use.
2) Data type conflict. Our resolution is to reserve the one of higher precision type or stronger expression. Data type conversion is needed.
3) Element substructure conflict. This conflict occurs when a leaf element is matched to an internal element. Our resolution is to reserve the internal element's substructure. In fact, this kind of conflict can be automatically resolved during the operation of algorithm in section 4.2.
4) Nested path structure conflict. This kind of conflict is the most complicated one. It arises from the inconsistency of two paths from currently matched elements and already matched ancestors elements in two schemas.

The nested path structure conflict can be subdivided to three types of conflicts:

1) Path in mediated schema contains nested element. See **Fig. 3**, there is an intermediated element labeled 'Jiaotong' between 'Shenghuo Peitao' and 'Gongjiao' the in mediated schema. This kind of conflict can be automatically resolved



▲Figure 3. Nested path conflict type one.

during the operation of algorithm in section 4.2.

2) Path in source schema contains nested element. See **Fig. 4**, there is an intermediated element labeled 'Jiaotong' between 'Shenghuo Peitao' and 'Gongjiao' the in mediated schema. If we do not take any measures, the integrated mediated schema will look like new mediated schema 1 because we cannot find a matching object for 'Jiaotong', it will be added to the mediated schema as the right-most child element of 'Shenghuo Peitao', which is a mistake. However, this kind of mistake can be detected later when it comes to 'Gongjiao', its original father element 'Jiaotong' in source schema now becomes its brother. Our solution is to cut down the connection between 'Gongjiao' and 'Shenghuo Peitao' in the mediated schema and add a connection between 'Gongjiao' and its new father element 'Jiaotong'. Finally, we get the correct mediated schema, the new mediated schema 2.

3) There are nested elements in two schemas and the two nested elements are falsely not matched. See Fig. 2, 'Goufang Jingjiren' and 'Zheye jingli' are dissimilar in string, so they are falsely not matched and 'Zheye jingli' is placed into the mediated schema as brother element of 'Goufang Jingjiren'. This forms a new mediated schema 1, which is a mistake. However, this kind of mistake can be detected later when it comes to element 'Mingzi'; its original father element in source schema now becomes its father's brother. Our solution is to cancel the previous matching decision of element 'Zheye jingli', and match it to the element 'Goufang Jingjiren' according to the matching result of their child elements. Finally, we will get the correct mediated schema, the new mediated schema 2.

5 Evaluation

5.1 Datasets

Experimental data is collected from multi-source heterogeneous data sets in second-hand housing domain, including second-hand housing information published by anjuke (http://www.anjuke.com), 5i5j (http://bj.5i5j.com), and lianjia (http://www.lianjia.com) in January 2015. Their XML schemas contain 51, 50 and 46 elements, respectively. We first do data integration on these schemas manually to get a standard result. Basic information about experimental data set is shown in **Table 1**.

5.2 Metrics and Baseline Method

We compare the matching result of our approach and the manual result, and primary evaluating metrics are precision rate (precision), recall rate (recall), and F-measure (F_measure). The equations used to calculate these are as follows. TP is the number of matching element pairs that exist in both results. FP is the number of matching element pairs that exists only in results of our approach. FN is the number of matching element pairs that exists only in the manual result.

▼Table 1. Basic information about data set

Number of source schemas	Total number of elements	Total number of instance data items	Total number of element pairs	Number of matched element pairs	Number of elements in manual mediated schema
3	147	30,121	7196	108	73

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (5)$$

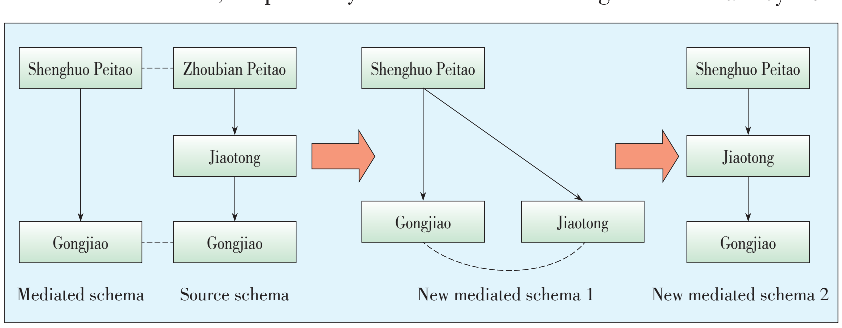
$$F\_measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

Performance oriented schema mediation (POSCHE) is the most similar related work with us, and we both do data integration in an incremental manner. Therefore, we choose POSCHE for comparison. In the meantime, whether there exists manual intervention in our framework imposes great impact on experimental result. Our framework makes decision on its own if there is no human intervention. Also, we conduct experiments on the effect of human intervention extent on the accuracy of our approach.

5.3 Experimental Result

5.3.1 Influence of Manual Intervention

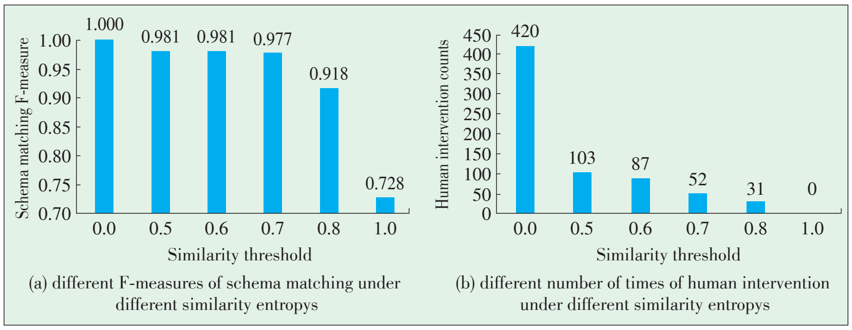
We set the threshold of similarity entropy, denoted T, to 0.0, 0.5, 0.6, 0.7, 0.8 and 1.0 to conduct six experiments, respectively. We record the number of expert queries and F-measures for every experiment. T=0.0 represents that decision is made all by human while T=1.0 represents decision is made all by machine. Experimental result is shown in **Fig. 5**. We can see that more human intervention corresponds to better F-measure. In overall consideration of human degree of effort and approach performance, we set T to 0.7, under which little human effort is paid but performance is much improved.



▲Figure 4. Nested path conflict type two.

5.3.2 Performance of Schema Matching

We compare the performance of POSCHE, our framework without human intervention (T=1.0),



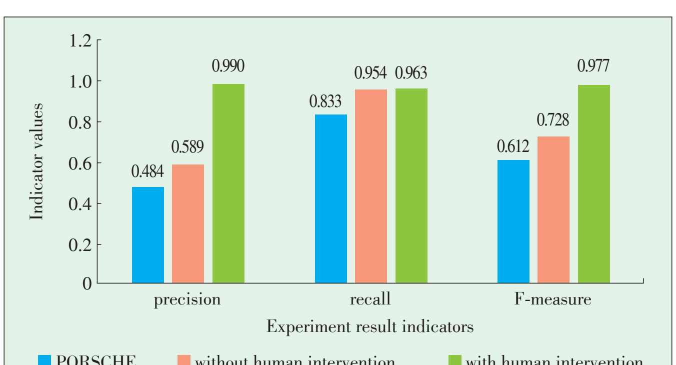
▲**Figure 5. Human intervention experiment.**

our framework with human intervention (T=0.7) in three experiments. Evaluating indicators are precision rate, recall rate and F-measure (**Fig. 6**).

Our framework with human intervention is better than our framework without human intervention in terms of all three indicators. Our framework without human intervention is better than POSCHE. In the experiment using our framework with human intervention, only 52 questions were posed to experts, which account for only 0.7% of the whole search space (totally 7196 pairs) and 12.4% of all questions (totally 420 questions are needed if all handled by human), but the F-measure reached up to 97.7%.

6 Conclusion

Aiming at the problems in domain of human-centric services, we propose a novel approach of schema integration with data from domain of human-centric services. In our approach, we use a mediated schema to help quickly integrate multiple schemas. Every schema is matched and integrated to the mediated schema only once (i.e., one iteration) and the mediated schema is updated and extended after each iteration. During each iteration, a depth-first search algorithm is used to control element matching and integration order. Five matchers which utilize both schema metadata and instance data are combined to complete schema matching. We introduced a similarity entropy based interactive method of human intervention controlling to make matching results more precise. After schema matching, a set of conflict resolution strategy is used to solve all kinds of complex conflicts and then form a better and more complete new mediated schema. We finally use real second-hand housing data from the internet to design and conduct experiments. The results show that our approach performs very well and requires very little human intervention.



▲**Figure 6. Schema matching experiment.**

A limitation of our approach is that although the degree of human intervention can be minimized, performance is not satisfactory if there is no human intervention. In the future, we will further study automatic schema matching and integration algorithms in order to improve matching performance with no human intervention. Another future work is to study entity matching work which is another important technique to fully accomplish the task of forming a complete data set in smart city.

References

[1] J. Madhavan, P. A. Bernstein, and E. Rahm, "Generic schema matching with cupid," Microsoft Research, Microsoft Corporation, Tech. Rep. MSR-TR-2001-58, Aug. 2001.
[2] D. Aumueller, H.-H. Do, S. Massmann, and E. Rahm, "Schema and ontology matching with COMA++," in *ACM SIGMOD International Conference on Management of Data*, Baltimore, USA, Jun. 2005. doi: 10.1145/1066157.1066283.
[3] N. F. Noy and M. A. Musen, "Anchor-PROMPT: using non-local context for semantic matching," in *Workshop on Ontologies and Information Sharing at IJCAI*, Aug. 2001, Seattle, USA.
[4] S. Madria, K. Passi, and S. Bhowmick, "An XML schema integration and query mechanism system," *Data & Knowledge Engineering*, vol. 65, no. 2, pp. 266–303, May 2008.
[5] R. A. Pottinger and P. A. Bernstein, "Merging models based on given correspondences," in *29th International Conference on Very Large Data Bases*, Berlin, Germany, Sept. 2003, pp. 862–873.
[6] P. Shvaiko and J. Euzenat, "A survey of schema-based matching approaches," *Journal on Data Semantics IV*, vo. 3730, pp. 146–171, 2005. doi: 10.1007/11603412_5.
[7] P. Bouquet, L. Serafini, and S. Zanobini, "Semantic coordination: a new approach and an application," in *Second International Semantic Web Conference*, Sanibel Island, USA, Oct. 2003, pp. 130–145. doi: 10.1007/978-3-540-39718-2_9.
[8] F. Giunchiglia and M. Yatskevich, "Element level semantic matching," in *Meaning Coordination and Negotiation Workshop at ISWC*, Hiroshima, Japan, Nov. 2004.
[9] B. He, K. C.-C. Chang, and J. Han, "Discovering complex matchings across web query interfaces: a correlation mining approach," in *Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Seattle, USA, Aug. 2004. doi: 10.1145/1014052.1014071.
[10] W. Su, J. Wang, and F. Lochovsky, "Holistic query interface matching using parallel schema matching," in *22nd International Conference on Data Engineering*, Atlanta, USA, Apr. 2006, pp. 122–125. doi: 10.1109/ICDE.2006.77.
[11] K. Zhang and D. Shasha, "Simple fast algorithms for the editing distance between trees and related problems," *SIAM Journal on Computing*, vol. 18, no. 6, pp. 1245–1262, Dec. 1989. doi: 10.1137/0218082.
[12] D. Shasha, J. T. L. Wang, and R. Giugno, "Algorithmics and applications of tree and graph searching," in *Twenty-First ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, Madison, USA, 2002, pp. 39–52. doi: 10.1145/543613.543620.
[13] C. Liu, J. Wang, and Y. Han, "Mashroom+: an interactive data mashup approach with uncertainty handling," *Journal of Grid Computing*, vol. 12, no. 2, pp. 221–244, Jun. 2014. doi: 10.1007/s10723-013-9280-5.
[14] W.-S. Li, C. Clifton, and S.-Y. Liu, "Database integration using neural networks: implementation and experiences," *Knowledge and Information Systems*, vol. 2, no. 1, pp. 73–96, Mar. 2000. doi: 10.1007/s101150050004.
[15] S. Massmann and E. Rahm, "Evaluating instance-based matching of web directories," in *11th International Workshop on the Web and Databases*, Vancouver,

Canada, Jun. 2008.
[16] J. Fan, M. Lu, B. C. Ooi, *et al.*, "A hybrid machine-crowdsourcing system for matching web tables," in *IEEE 30th International Conference on Data Engineering*, Chicago, USA, 2014, pp. 976–987. doi: 10.1109/ICDE.2014.6816716.
[17] H.-Q. Nguyen, D. Taniar, W. Rahayu, and K. Nguyen, "Double-layered schema integration of heterogeneous XML sources," *Journal of Systems and Software*, vol. 84, no. 1, pp. 63–76, Jan. 2011. doi: 10.1016/j.jss.2010.07.055.
[18] A. Aboulnaga and K. el Gebaly, "μBE: user guided source selection and schema mediation for internet scale data integration," in *IEEE 23rd International Conference on Data Engineering*, Istanbul, Turkey, Apr. 2007, pp. 186–195. doi: 10.1109/ICDE.2007.367864.
[19] K. Saleem, Z. Bellahsene, and E. Hunt, "PORSCHE: performance oriented schema mediation," *Information Systems*, vol. 33, no. 7, pp. 637–657, Nov. 2008. doi: 10.1016/j.is.2008.01.010.
[20] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *International Conference on Learning Representations*, Scottsdale, USA, May 2013.

Manuscript received: 2015-08-26

Biographies

Ding Xia (847525974@qq.com) is a postgraduate of Department of Information Science and Technology, Peking University, China. His research interests including ubiquitous computing.

Da Cui (443021181@qq.com) is a postgraduate of Department of Information Science and Technology, Peking University, China. His research interests including ubiquitous computing.

Jiangtao Wang (jiangtaowang@pku.edu.cn) is a postdoc researcher of Department of Information Science and Technology, Peking University, China. His research interests including mobile crowdsensing and ubiquitous computing.

Yasha Wang (wangyasha@pku.edu.cn) , PhD, is a professor of National Engineering and Research Center of Software Engineering, Peking University, China. His research interests including software reuse, data analytics, ubiquitous computing.

DOI: 10.3969/j. issn. 1673-5188. 2015. 04. 005
http://www.cnki.net/kcms/detail/34.1294.TN.20151125.1734.002.html, published online November 25, 2015

Top-Level Design of Smart City Based on "Integration of Four Plans"

Jianbo Cheng and Peng Sun
(ZTE Corporation, Shanghai 201203, China)

Abstract

A smart city provides a new idea and model for urban construction, management, and development. This article proposes the concepts and methodology for top-level design of smart cities based on the "Integration of Four Plans", and the planning process and systems for implementing the top-level design of smart cities. This article discusses, from the perspective of a city, how to optimize resource allocation, coordinate the development of urban economy, society, resources, environment, and people's livelihood, and map out the blueprints for healthy and sustainable development of a smart city.

Keywords

smart city; top-level design, integration of Four Plans, ICT Benefiting People

1 Top-Level Design Guiding Smart City Construction

A city is composed of many ecosystems, which is called a "System of System". From the perspective of urban structure, a city is a combination of interactive elements, such as the economic structure, social structure, and spatial structure. From a technical perspective, a city is a complex system composed of multi-field, multi-class, multi-level, and multi-directional heterogeneous sub-systems for collaboratively processing different types of massive data in different areas. A city has the following features:

- Rapid changes in demands
- Technological diversity
- Complex internal structure
- Frequent external interactions
- Coexistence of old and new systems
- Needs for continuous improvement.

Evolving information technologies and requirements for urbanization are jointly spurring the development of smart cities. A smart city is a new idea and model for urban construction, management, and development to improve urban management efficiency, facilitate public life, and promote technological innovation using information and communications technologies (ICTs), such as high-speed Internet, big data, Internet of Things (IoT), and cloud computing. Smart cities have been witnessing rapid development in China since 2012. Statistics shows that national ministries have approved a total of 529 pilot smart cities in Eastern, Central and Western China [1].

However, in the existing administrative systems, due to a lack of experience in smart city construction, different understanding and organization of smart cities across regions, and absence of state-level standards and evaluation systems, if without overall planning and guidance, many existing and new issues, such as lack of coordination, information silos, the same picture of different cities, and low degree of smartness, may arise again during the implementation process, increasing the risks of failure.

In respect of the complexity of smart city construction, this article proposes the top-level design of smart city as well as its implications, scope of planning, key planning contents, implementation strategies, and typical cases, to provide a feasible way for smart city construction through coordination of all levels and elements.

Top-level design is a concept of system engineering, and embodies the "overall idea" of a project. In order to complete a large project, overall planning from a global perspective with "common ideas, functional coordination, unified structure, resource sharing, and standardized components" is needed. In addition, a smart city is a long-term and sustainable systems project to reform the urban management ideas and models of governments, and cannot be promoted effectively without an appropriate top-level design [2].

When the smart city construction is implemented all over China, the following common problems arise from the practice during the top-level design process:

1) Focus on technologies rather than services

Top-level design was driven by technologies, and is merely in pursuit of application of hi-techs, such as the IoT, cloud computing, and mobile Internet. Without considering in-depth integration of ICTs with urban function modules or elimination of deep-seated contradictions, top-level design was always unattractive to the public. A smart city does not depend on advanced technologies but should properly tackle existing problems.

2) Focus on construction rather than management

The current smart city planning and design basically put emphasis on technical architecture and specific applications, but seldom take into account the closely-related operation management mechanisms, service models, policies, and measures, for example, how to build the libraries for information resource sharing, examination and approval lists, and public service lists, and how to implement dynamic management. Due to a lack of national standards, regulations, and procedures, and great differences in local circumstances, these are often difficult to achieve, and are usually the weaknesses of IT professionals. However, these are the basis and guarantee for achieving the goals of smart cities.

3) Focus on planning rather than implementation

Although a series of projects are planned in top-level design, local governments fail to implement these projects due to insufficient funding. The governments do not have constructive ideas on how to innovate in business models, attract social investments, involve financial institutions in smart city construction, and promote the development of the industry chain. Neither do the governments make clear plans for opening up the charged information resources of governmental departments or motivate the community through purchase of services, thereby prohibiting sustainable development of smart city construction.

The top-level design of smart cities is an innovative activity to promote urban development through research on urban planning, current urban situation, and urban economic and social development planning.

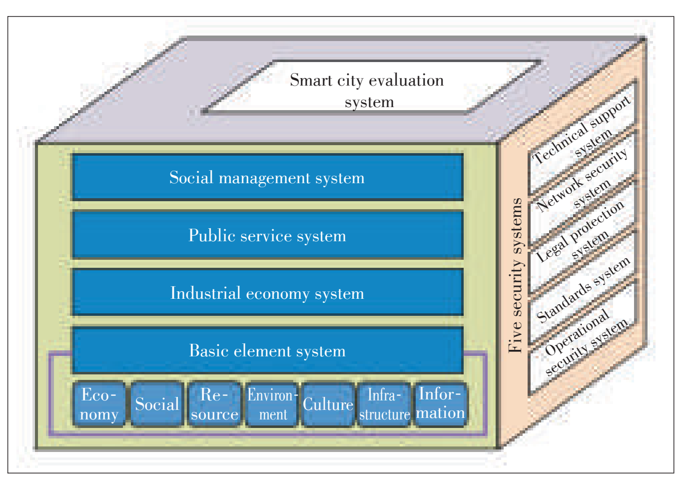
1) Future-proof plans, and practical and forward-looking blueprints can be formulated through the top-level design of smart city based on regional positioning and advantages of cities, and the requirements and trends for social, economic, and technological development in quite a long period of time.
2) Overall planning should be conducted from the perspectives of the demand framework, technological implementation, investment and financing modes, and industrial support, to improve organizational collaboration and achieve the highest efficiency and optimal resource allocation at the lowest costs.
3) Innovations should be made in IT project implementation, and the mechanisms and modes are designed deliberately to involve more enterprises in smart city construction, leveraging more social investments at a small amount of governmental expenses.
4) The core demands of governments, the public, and enterprises should be considered systematically to build a livable and sustainable smart city, and a convenient, fair, inclusive, high-quality, and efficient public service system, so that the public can enjoy the convenience and comfort of the smart city.

2 Implications of Top-Level Design of Smart City

Based on scientific methodologies, the top-level design of smart cities involves strategic planning and overall design of a city to optimize resource allocation, coordinate the development of urban economy, society, resources, environment, and people's livelihood, and map out the blueprints to guarantee healthy and sustainable development of a smart city. The top-level design of a smart city is not only for the city itself, but also for urban development and public experience. Urban development involves social, economic, and cultural value aspects, while public experience refers to an IT-based living environment that brings perceivable welfare to people.

From a different dimension, the top-level design framework of a smart city includes the basic element system, industrial economy system, public service system, social management system, evaluation system, as well as five security systems (technical support, security, legal protection, standards, and operational security), as shown in **Fig. 1**.

The basic element system, which is the foundation for operation and development of a smart city, improves efficiency, and facilitates innovation and coordination between urban operation lines and functional elements, including the economic, social, resources, environmental, cultural, infrastructure, and IT fields. The industrial economy system transforms and improves traditional industries through the new-generation information technologies to accelerate transformation and upgrade, and promote the rapid development of emerging strategic industries. The public service system is a comprehensive service system involving the public and enterprises to meet the main needs of the city. The social management system, which is at the core of the city's comprehensive regulations and management, imple-



▲**Figure 1. Top-level design framework of smart city.**

ments intelligent analysis and control of urban operations on the basis of efficient operations of the urban resource element system. The evaluation system guarantees that the city develops towards people - orientation, smart people, intelligent things, and optimal development of economic and social activities. The five security systems support and guarantee the operations, management, and services throughout the life cycle of the smart city.

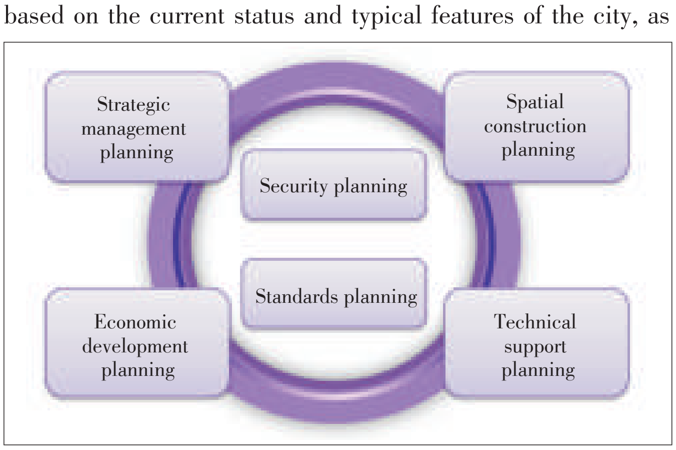
Having made lots of practices, ZTE proposes the top-level design methodology with "Integration of Four Plans". Namely, strategic management planning, spatial construction planning, economic development planning, and technical support planning are indispensable and their collaboration should be considered in the top-level design process of smart cities.

Based on strategic management planning, with the goal of economic development planning, guaranteed by technical planning, with overall spatial deployment, equal basic public services should be actively promoted to avoid redundant construction, and promote rational distribution and full use of resources to establish a unified smart city planning system with functional complementation. On this basis, a complete top-level design of smart cities can be formulated by involving security and standards planning (**Fig. 2**).

The key to top-level design is to develop a complete and feasible smart system. In this system, all parties can independently construct the system to build a sustainable framework for a smart city with a clear division of labor and joint efforts. Through the top-level design of smart cities, with unified strategic planning for urban development, the city positioning, and the contents, modes, and overall design of smart city construction should be identified to avoid decentralized investment and redundant construction, providing a feasible way of resolving the difficulties in the development of smart cities.

1) Strategic management planning

Strategic management planning, with the highest priority among the four plans, identifies urban development goals, regional planning, policy environment, and security systems based on the current status and typical features of the city, as well as makes reversal correction to city positioning.



▲Figure 2. Top-level design of smart city based on "Integration of Four Plans".

2) Spatial construction planning

Spatial construction planning includes urban spatial planning, land use planning, and municipal facilities planning based on the city positioning. It is a complicated systematic project involving the deployment and arrangements for political, economic, cultural, social, and land resources. Spatial construction planning is closely associated with the smart city to improve urban planning, promote the optimal design of public facilities, and achieve scientific and accurate urban planning using big data.

3) Economic development planning

Economic development planning includes planning for industrial development, economic environment, investment and financing, and financial services based on city positioning, in order to ensure sustainable construction and operation of a smart city. For example, industrial transformation and upgrade should be facilitated by planning the development paths for key industries, proposing recommendations for investment and financing models, and integrating information technologies, to ensure sustainable development of the smart city.

4) Technical support planning

Technical support planning is to achieve the strategic goals of a smart city by means of IT, where the difficulties include interconnection between systems, information mining, and mass data processing. With the development of the new-generation information technologies, such as the IoT, cloud computing, big data, and mobile Internet, data becomes ubiquitous, and is accumulating at any time. How to eliminate the long-existing information silos and conduct data mining to benefit people, promote business, and improve governance is key to the smart city construction, and needs to rely on technical support planning [3].

5) Security planning

Security planning is to build a unified information security system that involves security management, technical defense, secure operation and maintenance (O&M) capabilities, and unified management and control of the physical layer, sensor, network transmission, data, application, and O&M risks, to ensure data security and secure Internet access for a smart city.

6) Standards planning

A smart city involves a large number of areas, processes, systems, interfaces, as well as managers, constructors, and maintainers, and therefore needs to establish unified technical standards and regulations based on a sophisticated system for efficient collaboration, and standardize smart city construction projects through a unified, open, and operable construction standards system.

3 Contents of Top-Level Design of Smart City

In accordance with the planning concept of "Integration of

Four Plans" for top-level design, strategic management planning should be developed by analyzing the demands for smart cities. Strategic management planning for a smart city includes the following contents:

1) Guiding principle

Everything develops by following related principles, and the same is true for smart city construction, which also needs an appropriate guiding principle. In August 2014, the *Guiding Opinions on Promoting Healthy Development of Smart City* was officially released, clearly putting forward the guiding principle for smart cities: In accordance with the general requirements for following an intensive, intelligent, green, low-carbon road of urbanization, the market should play a decisive role in resource allocation, and governmental guidance should be strengthened and improved for coordination of materials, information, and intellectual resources, to promote the application of the new-generation IT innovations, strengthen construction of intelligent urban management and service systems, proactively develop intelligent applications for public services, and enhance network security, thereby improving the city's comprehensive capacity and residents' happiness, and enhancing the quality and level of urbanization.

2) Development goals

Construction will be in disorder without clear goals. Therefore, the social, economic, and livelihood goals of smart city construction should be determined in accordance with the city's regional positioning, strengths, development opportunities, and industry features. Social goals include caring public services, rational social management, optimal resource allocation, environmental improvement, effective economic development, people's welfare, and harmonious social trends. Economic goals include healthy development of industrial economy, continuous cost reduction, and increasing efficiency and benefits. Livelihood goals are to provide people with better living and guarantee services, such as basic housing, community, and public education services.

3) Key tasks

With adequate research on and analysis of demands, in accordance with smart city construction and development goals, key tasks and projects should be determined, target tasks and responsibilities should be specified, and construction requirements, leaders in charge, leading departments and accountable persons, and assistance departments and accountable persons should be identified for all construction projects. Key tasks should be prioritized in the planning process, highlighting priorities and implementation methods. For example, priorities for IT applications shall be specified in urban planning and management, public services, social management, infrastructure, and industrial development, and the issues constraining urban development should be addressed to explore and promote the sound development of information networks, technological innovations, industrial transformation, and social applications.

4) Strategic steps.

In accordance with the top-level design, a smart city should be conducted step by step by tackling easier issues first, focusing on usability, and launching pilot projects, to build, use, summarize, adjust, and improve the smart city simultaneously. For example, infrastructure should be improved with the guidance of "Broadband China", and big data centers should be built with the guidance of "ICT Benefiting People", including four basic databases for natural persons, legal persons, macroeconomics, and spatial geography. Furthermore, corresponding subsystems, such as smart transportation and smart tourism, should be built to meet different local needs. The smart city construction is in progress, and will never end.

Upon the completion of strategic management planning, further measures should be taken for the collection, integration, storage, mining, and sharing of information resources.

First, a technical support system should be developed for the smart city. A smart city has a long-term development process, and therefore needs a scalable, flexible, and evolving technical system. This system coordinates all project levels and elements from a global perspective, analyzes the current status, target architecture design, interconnection specifications, and comprehensive construction constraints in respect of services, data, systems, and technologies, and designs implementation path for the target architecture [3].

Second, a service application system should be developed for the smart city. With the goal of fully supporting government functions, the service application system is the infrastructure for planning and designing service functions of a smart city, breaking the limitations of specific department functions, and reviewing the services in respect of serving the public, in order to adapt to the changes in urban management, economic development, people's livelihood, and public services. This system better serves the general public, governmental departments, and enterprises by building related service application systems, such as urban planning, construction, operations, and management services.

Third, a policy and standards system should be developed for the smart city. Currently, the level of standardization has become an essential element of core competitiveness of all countries and regions in the world. In the top-level planning stage of smart cities, great importance should be attached to standardization. There must be guides for planning smart cities, standards for building shared platforms, and references for collaborative application and development, to achieve orderly, high-quality, and significant development.

Fourth, an evaluation system should be developed for the smart city. The evaluation system should be established for scientific measurement of smart city construction, and specifying and setting indicators for the smart city, to ensure rational urban investment, efficient management, livable environment, and happiness of residents. The evaluation system identifies the construction goals and standards of the smart city, to help smart city administrators make objective judgment on the construction process and results, and qualitative and scientific evaluation of the construction achievements. This system has become an effective means for assessing smart city construction, and is of great significance and value for solving these problems. In accordance with the upcoming state-level evaluation indicators for a smart city, based on the actual conditions, detailed evaluation systems can be developed for all cities [1].

Fifth, an investment and financing system should be developed for the capital industry of a smart city. Smart city construction is a complicated systematic project that has wide coverage, and needs huge capital investments. It cannot be effectively supported by relying on only the public expenditure of a nation. It is an inevitable choice to meet the funding needs of smart city construction through various flexible investment and financing models. For example, the public–private partnership (PPP) model can be used to introduce market mechanisms into smart city construction projects. In this way, additional funding needs can be fulfilled by enterprises and financial institutions, achieving the goals of shared and win-win project construction involving multiple parties. This will become an inevitable choice for investment and financing models of smart city construction projects in China.

4 Cases of Top-Level Design of Smart City

Y City, located to the south of the Yangtze River and to the north of the Changsha-Zhuzhou-Xiangtan City Group, has rich natural and tourism resources, and is well-known as "a land flowing with milk and honey", the "Land of Black Tea", the "Land of Ramie", the "Land of Nan Bamboo", and the "Land of Culture" in China. With distinctive geographical advantages and rich element resources, Y City has witnessed the rapid development of various undertakings, increasing economic strength, improving information infrastructure, rapid development of the information industry, and better environment for IT-based development in recent years, laying a solid foundation for smart city construction. However, Y City also faces a series of problems and challenges:

- Closed platforms and imbalanced development result in information silos, and information sharing and opening up lack intrinsic motivation or external regulation mechanisms.
- Overall planning and construction should be further improved due to poor IT-based coordination.
- IT infrastructure does not meet the development requirements, with a low level of intensive development and weak information security capabilities.
- Due to low investment in IT, Y City suffers a severe shortage of talents.

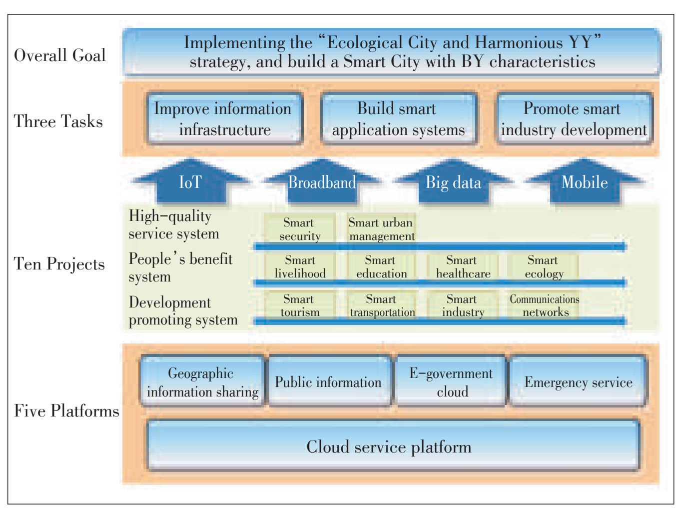
Based on its actual conditions and geographical advantages, with adequate research and demand analysis, Y City proposes the purposes of providing high-quality services, benefiting people, and promoting development, and the goal of enhancing the city's core competitiveness for smart city construction through "strategic management planning". Y City promotes the improvement and optimization of public infrastructure through spatial construction planning, determines to promote the transformation and upgrading of traditional industries and strategic development of the IT industry through economic development planning, and identifies to build an urban platforms for public services to achieve city-level information gathering and resource integration through technical support planning. In addition, Y city guarantees smooth implementation of the smart city through security planning and standards planning.

Y City uses the top-level design with "Integration of Four Plans", and formulates the framework of smart city construction, including three major tasks, ten major projects, and five key platforms (**Fig. 3**).

Based on the development strategy of "Ecological City, and Harmonious YY", in accordance with the general requirements for following an intensive, intelligent, green, low-carbon road of urbanization, with the goal of enhancing the city's core competitiveness, this top-level design promotes the application of the new-generation IT innovations, and boosts economy restructuring and social management innovation, providing important support and laying a solid foundation for build livable Y City with economic prosperity, social civilization, and beautiful environment.

To ensure the smooth implementation of the smart city, Y City starts from improving information infrastructure, building smart application systems, and promoting industrial development to lay a solid foundation for smart city construction, achieve standardized, accurate, and intelligent urban construction and management, and cultivate and develop emerging strategic industries, creating new economic growth poles.

To effectively promote citywide public resource sharing, achieve collaborative and efficient operations of people, logistics, information flow, and capital flow, and improve its operational efficiency and public services, by focusing on providing high-quality services, benefiting people, and promoting devel-



▲**Figure 3. Case of top-level design of a smart city.**

opment, Y City makes overall planning of ten projects, including smart security, smart urban management, smart livelihood, smart education, smart healthcare, smart ecology, smart tourism, smart transportation, smart industry, and smart communications networks.

To promote intensive construction of the smart city, and rational use of existing resources to avoid redundant construction, ZTE provides a full range of support for key projects of Y City through overall planning of five major platforms, such as the cloud service platform, geographic information sharing platform, the public information platform, e-government cloud platform, and emergency service platform, thereby raising the overall level of the smart city.

Through the top-level design of a smart city, starting from big data mining and applications, Y City implements online processing of all governmental activities, formulates citywide information resource data standards, launches information resource update, exchange, sharing, and opening-up policies and measures, and improves the governmental information resource catalog system and exchange system. In addition, Y City uses big data technologies for multi-dimensional analysis and judgment of economic and social development, to fully explore and enhance the value of data resources, and provide support for scientific prediction, correct decision-making, and quick processing. By improving appropriate security mechanisms for smart city development, such as accelerating urban information infrastructure and strengthening information security, Y City focuses on the construction of major smart application projects, and makes great breakthroughs in urban management and public services.

5 Summary

In general, smart city construction in China is still in the early and the exploratory stage. What is a smart city? How to build a smart city? Due to different understanding of these fundamental questions, many cities may encounter such problems as information silos, same images of different cities, and low degree of smartness during the smart city development. The root causes include management without collaboration due to lack of top-level design and overall planning, lack of innovations in systems and mechanisms, and potential network security risks. Some places even have the signs of muddy thinking and blind construction, and are therefore in urgent need of strengthening guidance. Based on overall needs for smart city development, through the top-level design concepts and methodology of smart city with "Integration of Four Plans", we should integrate and coordinate all resources from top to bottom, identify the relationships between all tasks, and design the implementation path systematically to ensure healthy and sustainable development of the smart city.

References

[1] B. Qiu, *Research Report on Development of Smart Cities in China (2012–2013)*. Beijing, China: China Architecture and Building Press, 2013.

[2] D. Cheng, *Introduction to the Top Design of Smart City*. Beijing, China: Science Press, 2012.

[3] J. Wang, C. Li, Z. Xiong, and Z. Shan, "survey of data-centric smart city," *Journal of Computer Research and Development*, vol. 51, no. 2, pp. 239–259, Feb. 2014.

[4] Y. Wang, "Smart city top-down design," *ZTE Technology Journal*, vol. 20, no. 4, Aug. 2014.

Manuscript received: 2015-09-08

Biographies

Jianbo Cheng (cheng.jianbo6@zte.com.cn) received his master's degree from Xi'an Jiaotong University, China. He is the chief engineer of the Smart City College of ZTE. He is responsible for the top-level design, consulting and planning for solution design, and project implementation of smart cities.

Peng Sun (sun.peng@zte.com.cn) received his master's degree from Southeast University, China. He is the chief engineer of the Industry Solution Department, Government & Enterprise Business Division of ZTE, and a standing committee member of ZTE Technology Expert Committee. He is responsible for strategic planning and solution architecture design. He led the development of ZTE's overall architecture of smart cities, and has obtained 24 invention patent licenses as the first inventor.

DOI: 10.3969/j. issn. 1673-5188. 2015. 04. 006
http://www.cnki.net/kcms/detail/34.1294.TN.20151204.1047.002.html, published online December 4, 2015

Smart City Development in China: One City One Policy

Biyu Wan[1], Rong Ma[2], Weiru Zhou[1], and Guoqiang Zhang[1]
(1. National Smart City Joint Lab, CSUS, MOHURD, Beijing 100835 China;
2. iSoftStone Information Technology (Group) Co., Ltd, Beijing 100193, China)

Abstract

China is in a process of urbanization and is aiming at a type of people-centered urbanization. The main purpose of developing a "smart city" is to help this type urbanization and to serve the people of the city. From 2012 to 2015, China has chosen more than 300 cities or towns to be national pilot "smart cities." These pilot smart cities are located in more than 30 provinces around China, which differ greatly in thousands ways. So we advocated "One City One Policy". In 2012, MOHURD announced 90 cities as first batch of pilot smart cities. After three years, some pilot cities achieved great progress. This paper introduces five example cities (including town, district) as five different models of China's smart city development. They are: Guilin city; Yunlong demonstration zone; Panyu District; Yangling Agricultural Hi-tech Industries Demonstration Zone; Lecong town. This paper also introduces our standardization work on smart city field at present.

Keywords

urbanization; MOHURD; pilot smart city; smart tourism; smart city indicator system

1 Introduction

China's smart city development officially started in 2012, when the Ministry of Housing and Urban-Rural Development (MOHURD) announced 90 cities or towns to be the first batch of pilot smart cities. In April 2015, MOHURD announced the third batch. Now, MOHURD and the Ministry of Science and Technology are promoting the smart city pilots to work together, and in total, the number of three batch of national "pilot smart city" reaches 300. [1]

The overall idea of the national smart city pilot project is to start with solving the actual problems of a city or town through the comprehensive application of modern science and technology in order to add smartness to a city. The purpose of smart city development includes: urban planning and management of cities/towns/districts, the allocation of urban resources, creating liveable environment, and using cultural heritage and innovation. In other words, the development of smart cities aims to promote the happiness of citizens and focus on urban sustainable development.

Therefore, in the process of development of smart city, MOHURD asks all pilot cities and towns to insist on an issue-oriented, demand-oriented and goal-oriented approach. This means one city has to work out a comprehensive plan and follow five guide points from MOHURD. The guides are focusing on: "one city one policy, people-centered strategy, city-industry integration, urban-management system innovation; market-leading and optimizing resource allocation.

2 Five Pilot Smart Cities in China

After three years of development on smart city practice, some good results have been presented. Here, cases of five cities (towns) are introduced as models of China's smart city development. They are Guilin City, Yunlong Demonstration Zone, Panyu District, Yangling Agricultural Hi-Tech Industries Demonstration Zone, and Lecong Town.

2.1 Smart Tourism in Guilin

Guilin is a prefecture-level city in the northeast of the Guangxi Zhuang Autonomous Region, China. It is situated on the west bank of the Li River, bordering Hunan to the north. The name Guilin means "Forest of Sweet Osmanthus," owing to the large number of fragrant Sweet Osmanthus trees located in the city. The city has long been renowned for its karst mountains and is one of China's most popular tourist destinations. At the end of 2014, Guilin city's permanent resident population was 4.91 million—2.24 million urban population and 2.67 million rural population. Guilin's urbanization rate is 45.6%. The natural population growth rate is 6.34%.[2]

Guilin as an international famous tourist city is chosen by the State Council of China as one state-level tourism reform pilot area. In November 2012, the National Development and Reform Commission (NDRC) agreed "Guilin international tourism destination construction development plan", so the development of Guilin tourism is also a national strategy. Therefore, the Guilin municipal government attaches great importance to the “smart tourism” as a new breakthrough in transformation and upgrading of tourism, also as a new area to promote other industries of Guilin. **Fig. 1** shows demonstration spots of the smart tourism in Guilin.

Smart tourism involves using advanced technologies such as cloud computing, Internet of things, mobile communications, intelligent application, high performance information processing, data mining, and other ICT technology, to build a smarter application system for visitors, travel agencies, departments of Tourism Administration.

Guilin is building its smart tourism framework as “one platform, four systems, two demonstration projects, and eight series of tourism products”. In details, one platform is the Guilin public travel information service platform; four systems of smart tourism include service system, management system, marketing system, and tourism enterprises system; two demonstration projects are “one characteristic small town” and “one characteristic low carbon small town”; eight series of tourism products are water sightseeing, leisure vacation tourism, historical and cultural tourism, national culture tourism, red tourism, ecological village tourism, outdoor sports tourism and romantic wedding tour.

2.2 Yunlong Demonstration Zone: The Road Map for Low-Carbon Society and Eco-City

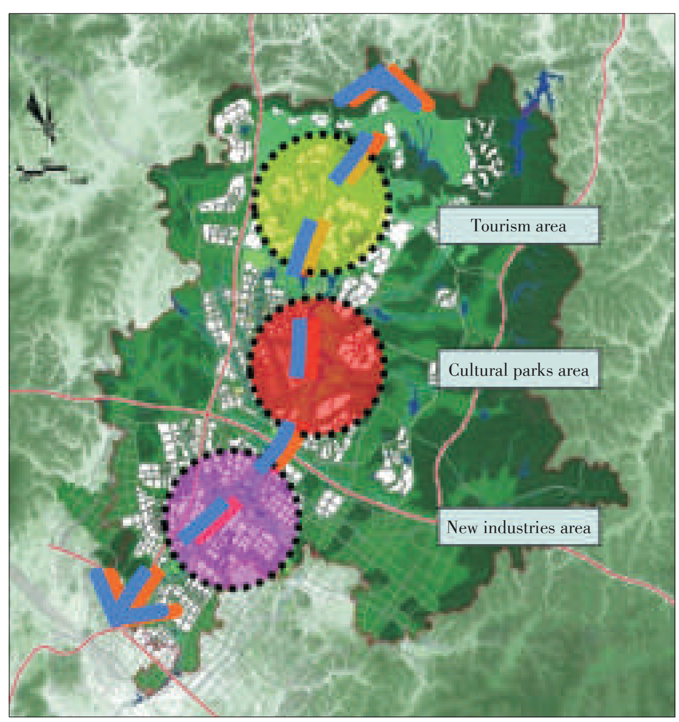
The Yunlong Demonstration Zone, established in 2009, is a new district of Zhuzhou City, Hunan Province. **Fig. 2** shows one image map of Yunlong Demonstration Zone. It is located in the northern part of Zhuzhou City, is the Fifth District of Zhuzhou, with a total area of 105.8 km^2 with a total population of 66,400 people, leading two towns, 22 villages and three community committees. [3] From the beginning, Yunlong summarized the road map for low-carbon society and eco-city, its urban design and planning aim to the two goals. And the Demonstration Zone is divided into three areas: tourist area, cultural park area and new industrial area as shown in **Fig. 3**.

▲**Figure 1. Demonstration spots of the smart tourism in Guilin.**

Yunlong made six detailed plans for its development: 1) green traffic, 2) low carbon energy, 3) ecological and environmental protection, 4) comprehensive utilization of water resources, 5) low carbon industrial, 6) urban management. Using the concept of smart city and technical tools, it will enhance the capability of urban infrastructure, command operations, ability to respond emergency events, better public service, etc.

In detail, the green traffic plan will develop public transport first, manage road usage, build public transport system and green road system. In water resources comprehensive utiliza-

▲**Figure 2. One image map of Yunlong Demonstration Zone.**



▲**Figure 3. Land planning map of Yunlong Demonstration Zone.**

tion, water conservation is as the core; the plan will promote the optimization of water resources allocation and recycling, strengthen the water ecological restoration and reconstruction, rationalize the use of rainwater, strengthen the surface water conservation, to build a safe, efficient, sustainable water environment. In land use, Yunlong will pay attention to developing towns and villages, reserving agricultural land at rural areas, and maintaining high-density development at urban areas so that the relative proportion of urban construction land, ecological land, and agricultural land use will be 1:1:1.

Finally, Yunlong new area is trying to use of build-operate-transfer (BOT), public-private partnership (PPP), transfer-operate-transfer (TOT), building-transfer (BT) and other new financing model, to attract foreign capital and private capital, promote the construction of infrastructure.

2.3 Smart Public Service at Panyu District

Panyu is a district of Guangzhou in southern China. It was formerly a county-level city before it became a district of Guangzhou. The name of Panyu dates back to the conquest of Guangdong by Qin Shi Huang. It was the old name of present-day Guangzhou. The district covers an area of about 661.88 km^2. It is consisted of six towns, 10 streets, total of 177 administrative villages, and 84 community committees. [4]

Panyu District has maintained a rapid economic development and the size of the population is also growing rapidly. Now the actual resident population is 3 million, far more than the population that has household registration. At present, one unresolved problem is that the effective supply of public service is not good enough to satisfy the public, and this problem is becoming increasingly prominent.

Therefore, Panyu District's goal in terms of smart city development is to promote people's livelihoods and provide better information services. Panyu plans to implement four projects related to people's livelihood: 1) file management plan, 2) livelihood service cards, 3) service points, 4) livelihood services. In particular, is to promote the "cloud services + smart service card + smart App" plan to push construction of a comprehensive and highly integrated intelligent ICT, providing the public with a number of smart services, such as emergency services, community services. **Fig. 4** shows the public service hall of Panyu government. For example, in 2013, the "Rental Housing Safety Management" worked well and is in charge of 643,000 rental housing sets. The coverage rate of 94.4%; and the "Citizen Hotline of the Panyu District" got 140,036 complaints or comments, 138,880 cases received satisfied reply, wind-up-case rate was 99.17%.

2.4 Smart Agriculture at Yangling Agricultural Hi-Tech Industries Demonstration Zone

Yangling Agricultural Hi-Tech Industries Demonstration Zone is located in Yangling District, Xianyang, Shaanxi. It is under the direct governance of Shaanxi Province, and is the only such zone within the People's Republic of China. It was created on 29 July 1997. This district has the planning area of 22.12 km^2, and its population is 80,000 people. [5]

Yangling Demonstration Zone believes that it is very important for modern agriculture to control whole agricultural industry chain of quality and safety and to improve food safety production. One vegetable factory at Yangling is shown in **Fig. 5**.

Yangling Demonstration Zone affixed to two dimension code as identification to its all agricultural products. Consumers can scan the code only using a smart mobile phone, can easily query the time of sowing, picking time, fertilization time and fertilizer name, the pesticide name used, time and production manufacturers, etc. In this way, all these agricultural products have an "identity card". **Fig. 6** shows one customer checks his food's identity. And, four local standards have been worked out for safety of Yangling agricultural products. These are standard for the management of traceability, technical specification for products-examination, examination criteria, and certification inspection standard.

2.5 Smart City in Lecong Town: Focusing on Industries

Lecong Town is situated in the hinterland of the Pearl River

▲Figure 4. Public service hall of Panyu government.

▲Figure 5. One vegetable factory at Yangling.

Figure 6. ▶
One customer checks his food's identity.

Delta, the northwestward of Shunde and the south of the central urban area of Foshan. It is less than 30 km from Lecong to Guangzhou and more than 100 km to Hong Kong and Macau. The National Highway 325 runs through from the south to the north. Lecong town covers 78 km^2, owns five community committees, 19 villages, household registration population of 100,000 people, and resident population of 260,000 people. [6]

As a small town of only 78 km^2, Lecong gains its achievements being the world's largest furniture market, the country's largest iron and steel market, South China's largest trading market of plastic. With the acceleration of urbanization and serious market competition, Lecong is facing many difficulties, such as wrong land planning, low end market, environmental degradation, and public service challenges.

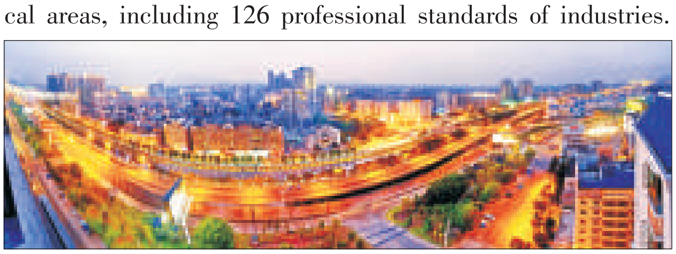
From 2012, Lecong local government started the development of Internet of things (IOT) industry as a new starting point, to increase investment, to start building "smart town" and to promote upgrade of Lecong's other traditional industries. Until now, IOT Industrial Park has been successfully opened, attracted more than 30 famous research institutions and enterprises into this Park, such as South Korea's Samsung data.

Lecong is one typical small town of China's first batch of smart pilot cities. "Smart Lecong" development is an important example in the eastern coastal areas different with other big cities. Through the IOT industry and other new ICT innovation with other industries, Lecong promotes the upgrading and transformation of traditional industries and the development of urban area. It has become a model of smart development in China's small cities and towns. **Fig. 7** shows a Lecong night scene.

3 Smart City Standardization and "One City One Policy"

Now, smart city development is in full swing, the demands of the standard system are also increasingly strong. At the end of 2012, MOHURD published "national smart city (town, district) pilot standards system (for trial)", a guide book for 90 pilot smart cities that year.

This indicator system contains six aspects of smart city: standards, infrastructure, construction and liveable, management and services, industry and economy, and security and maintenance. It can be divided into five levels and covers 18 technical areas, including 126 professional standards of industries.



▲**Figure 7. Lecong night scene.**

The smart city indicator system and 57 indexes for the pilot-project are shown in **Table 1**. Here are 57 indexes of these 6 aspects as third level. And these indexes are flexibly divided into required index, optional index and innovational index.

MOHURD's pilot smart cities are located in more than 30 provinces around China, and are different in many ways. For some small cities or underdeveloped areas, both the urban infrastructure and the ICT service could not take the same standards and criteria with developed areas. So China advocated "One City One Policy", will not judge all pilot cities with same standards and evaluation indicators in future.

National Smart City Joint Lab of Chinese Society for Urban Studies (CSUS) was set up in 2012. The joint lab consists of over 30 enterprises and research institutions, such as National Engineering Research Center for Information Security, Huawei, ZTE, Microsoft, IBM, MIT, HITACHI, Baidu, Xinhua News Agency Modern Express, Economic Daily, etc. The lab focus on the research and construction of all concerning with smart cities, such as ICT, public information platform, city security, and water engineering. Beginning with solving practical issues, the lab is endeavor on putting forward scientific, thorough and executable solutions for smart city construction, assisting pilot smart cities governance and offering consultation

▼**Table 1. Smart city indicator system**

Smart City Indicator System (Pilot, SCI)		
First Level	Second Level	Third Level
Guarantee system and infrastructure	Guarantee system	Planning outline, implementation scheme, organization guarantee, policy and regulation, funds guarantee, management of operation
	Network infrastructure	Wireless network, broadband network, next-generation broadcasting network
	Common platform and database	City common basic database, city common information platform, information security
Smart construction and livability	Administration of city construction	Urban and rural planning, digital city administration, construction market administration, house property administration, landscaping, historical and cultural preservation, building energy conservation, green building
	Promotion of city function	Water supply system, drainage system, water conservation application, gas system, garbage classification and disposal, heat supply system, lighting system, underground pipeline and spatial integrated administration
Smart Administration and Service	Governmental service	Decision support, information disclosure, online service, governmental service system
	Basic public service	Basic public education, labor employment service, social insurance, social service, medical and health service, public culture and sports, service for disabled, basic housing guarantee
	Special service	Smart transportation, smart energy, smart environmental protection, smart land administration, smart emergency, smart security, smart logistics, smart community, smart house and home, smart payment, smart finance
Smart industry and economy	Industry planning	Industry planning, innovation investment
	Industry upgrade	Industrial factors agglomeration, traditional industry transformation
	Development of emerging industry	High and new technology industry, modern service industry, other emerging industry

service.

Based on over 200 pilot smart cities practical experiences, the National Smart City Joint Lab is cooperated with its partners to continue standardization work of smart city. Recently, ISO/TC 268's TR37150 Chinese version has been published as one guidebook for city mayors and leaders.

4 Conclusion: Smart City and China's Urbanization

China is in a crucial stage of development. There is steady, coordinated progress in advancing the new type of industrialization, IT application, urbanization and agricultural modernization. However, there is a large urban-rural gap in development. Population in the central, western and northeastern regions accounts for over 60% of the national total. And per capita GDP has just exceeded US $5000. Narrowing the gap between urban and rural areas and between different regions will unleash huge potential for growth. China will implement a new type of people-centered urbanization and address the bifurcation between urban and rural areas. The government will increase support and use market tools to rebuild more rundown areas this year. Government will also promote the development of green industries, new energy, and energy-conserving and environment friendly technologies and products to foster new growth areas, and resolutely eliminate backward production facilities in this process to ease the resources and environmental constraints.

The international experience has proved that when urbanization reaches 50% to 70%, social problems tend to increase. Cities become sick; problem of unemployment, the gap between rich and poor, housing shortage, traffic congestion, energy shortage and environmental pollution, and other issues become serious. China's urbanization rate for the first time reached 50% in 2011. China's urbanization is in a critical period. To achieve people-centered urbanization, "smart city" helps. Smart city development requires to solve various urban diseases existed, and at the same time to prevent new problems or troubles in urban areas. The main purpose of developing a smart city is to serve the people better.

References

[1] B. WAN, "Smart city development needs one city one policy," *Smart City*, no. 8, pp. 64–65, Aug. 2014.
[2] *The Guilin Government* [Online]. Available: http://www.guilin.gov.cn
[3] *Yunlong Demonstration Zone* [Online]. Available: http://www.zzyunlong.gov.cn/Index.html
[4] *Panyu District* [Online]. Available: http://www.panyu.gov.cn
[5] *Yangling Agricultural Hi-Tech Industries Demonstration Zone* [Online]. Available: http://www.ylagri.gov.cn
[6] *Lecong Town* [Online]. Available: http://www.lidc.org.cn

Manuscript received: 2015-09-11

Biographies

Biyu Wan (wanbiyu@scitylab.org) is the CEO and chief scientist of National Smart City Joint Lab of Chinese Society for Urban Studies, and also is one senior expert of the Ministry of Housing and Urban-Rural Development. He received his PhD from Kobe University and worked in Japan for 10 years. He was awarded as "Excellent Young Scientist of Japan" when he worked in Kobe University. Now he is one of the most important leaders in the smart city field in China. He is also a vice chairman of ISO/TC 268 SC1 and expert of IEC SEG1 and ITU-T/SG20, leading smart city standardization work of China. He has visited more than 200 pilot cities and towns as MOHURD's expert. Many cities engaged him as their advisor or consultant. Dr. WAN has solid background in various fields including engineering, urban-planning, and the energy field. He has led various scientific projects and published a few books, many reports and articles. He has recently published the book "The Self-Sufficient City", introducing Spain's Barcelona city to Chinese readers.

Rong Ma (rongmac@isoftstone.com) received her PhD at Kobe University in 2003 and has been worked in Japan for 8 years. Now she is VP of iSoftStone Information Technology (Group) Co., Ltd at ZhongGuanChun Hi-Tech Park in Beijing. Her team of iSoftStone Group is working in the smart city field, helping more than 30 cities in China to develop smart city projects. Dr. MA is also a senior researcher of National Smart City Joint Lab of Chinese Society for Urban Studies. In particular, she is inductee of Beijing HaiJu-talent Plan (The Recruitment of Global Experts, or Thousand Talent Plan). She has solid background in IT technology, Engineering and Business. She is also a member of various scientific organizations and agencies. She has led various scientific projects and published a few books, more than 50 reports and articles in the last 10 years. Her book "One livable and sustainable city: Singapore" has recently published, introducing Singapore's experience in how to develop a livable and sustainable city.

Weiru Zhou (zhouweiru@scitylab.org) is the head of Theory and Standardization Department, Smart City Joint Lab of Chinese Society for Urban Studies. She is responsible for the international and Chinese smart city standards. She is a member of many international standard working groups, such as IEC/SEG1/WG2, ISO/TC268/SC1/AHG3, ITU-T/SG20. She is also an important leader of the China National Smart City Standardization Overall Group, which is responsible for organizing five China National Smart City standards. She has led various smart city projects.

Guoqiang Zhang (zhangguoqiang@scitylab.org) graduated from China University of Mining and Technology (Beijing) and now is working at National Smart City Joint Lab of Chinese Society for Urban Studies. He is mainly engaged in land management and application, remote sensing technology, theory and standard system research of smart city and planning advisory services of smart city. He has led various smart city projects.

DOI: 10.3969/j. issn. 1673-5188. 2015. 04. 007
http://www.cnki.net/kcms/detail/34.1294.TN.20151118.0733.002.html, published online November 18, 2015

Gateway Selection in MANET Based Integrated System: A Survey

Ye Miao, Zhili Sun, and Ning Wang
(Institute for Communication Systems (ICS), University of Surrey, Guildford, GU27XH, UK)

Abstract

Taking advantage of spontaneous and infrastructure-less behavior, a mobile ad hoc network (MANET) can be integrated with various networks to extend communication for different types of network services. In the integrated system, to provide inter-connection between different networks and provide data aggregation, the design of the gateway is vital. In some integrated networks with multiple gateways, proper gateway selection guarantees desirable QoS and optimization of network resource utilization. However, how to select gateway efficiently is still challenging in the integrated MANET systems with distributed behavior terminals and limited network resources. In this paper, we examine gateway selection problem from different aspects including information discovery behavior, selection criteria and decision-making entity. The benefits and drawbacks for each method are illustrated and compared. Based on the discussion, points of considerations are highlighted for future studies.

Keywords

gateway selection; MANET; integrated network; routing algorithms; QoS

1 Introduction

As mobile technology develops, mobile devices have become more powerful with high transmission speed, low power consumption and low cost. Hence, the paradigm of multihop ad hoc networking has become popular. An established mobile ad hoc network (MANET) can cooperate with various network systems to extend communication areas for applications and ensure better user experience. The application domains, e.g. public safety [1], intelligent transportation system, mesh community and mobile data off-loading [2], can benefit from such integration. One of the major challenges of the integrated networks is to serve large amount of traffic data from end devices with limited resources. Thus, in order to satisfy the Quality of Service (QoS) of the different applications and devices, proper and efficient mechanisms should be developed for suitable allocation of available network resources.

In an integrated system, a MANET node can communicate with external networks via gateways. A gateway node, with multiple interfaces and protocols, acts as interworking unit responsible for forwarding traffic between different types of network. To provide interoperability and end-to-end network connectivity for end devices, gateways need to be implemented with various networking functions that regulate data and control traffic. To optimize network resource utilization and achieve satisfied QoS, the data transmissions must be distributed and make best use of allocated resources accordingly which can be achieved by proper gateway selection mechanisms. However, due to the distributed behavior of mobile nodes and limited network resources, there are still some issues to be addresses when designing gateway selection mechanisms.

In this paper, we investigate the gateway-selection problem. The main contribution of this paper is to provide a survey of gateway selection mechanisms in a MANET based integrated networks. The previous studies are categorized in terms of different criteria, such as information discovery method, selection metric, and decision-making entity. We present explanations and comparison of the studies in each category, and specify some load-balancing based algorithms. With the development of hardware techniques, networks such as machine-to-machine network (M2M) and Internet of things (IoT) can also be integrated with MANETs for data aggregation. With a large number of mobile devices and enormous amounts of data, load-balancing is essential for alleviating traffic congestion and even traffic distribution in order to improve network performance. We also point out future study directions based on the surveyed papers and analysis.

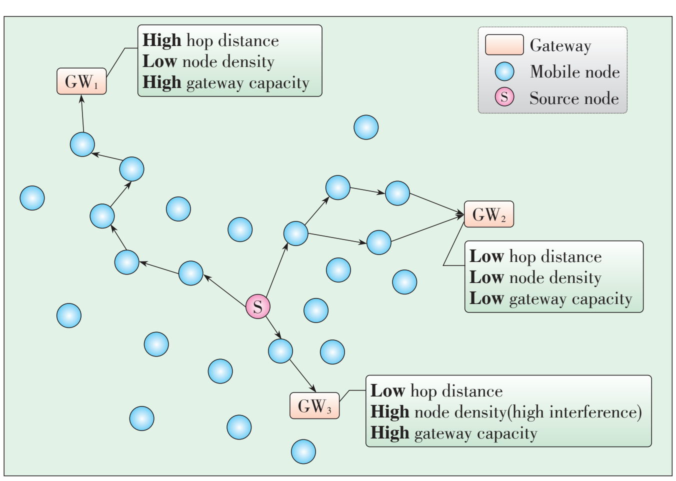
The rest of the paper is organized as follows: In section 2, we discuss and state properties of gateway selection problem. In section 3, the current gateway selection mechanisms in terms of different categories are reviewed. In section 4 we highlight the open research challenges and issues for future research. Finally, in section 5, we conclude the paper.

2 Gateway Selection Problem Statement and Discussion

In order to communicate with external networks, mobile nodes should target one or more trusted gateways to direct their traffic to. During this procedure, several aspects need to be considered. Firstly, the gateways are limited by processor capacity and outgoing transmission link capacity. Thus, the amount of traffic that goes through a gateway must not exceed the capacity of the gateways; otherwise, the data packets will be dropped. This may lead to more packets having to be retransmitted, longer delay, and poorer throughput. In addition, gateways may have different capacities of connections that are not known by the mobile nodes. This may cause unevenly utilization of gateways' resources. Furthermore, mobile nodes select gateways spontaneously without considering other nodes, it is possible for them to have the same preferable paths at the same time, which may cause congestions on wireless channel both in-network and around gateways.

In order to make best use of the network resources and avoid problems, it is essential to consider network performance during gateway and path selection, when there are multiple paths and gateways available in network. Gateway selection problems include 1) which gateway to select for each data packet flow, 2) how to route the flows through multi-hop network paths established between mobile nodes and selected gateways, and 3) increase network throughput and improve QoS in integrated system. Gateway selection problem can be broken into three steps: measurement, decision and execution [3]. The first step is path/gateway information discovery and collection. Information accuracy is guaranteed in this step. The second step is decision phase that concentrates on the algorithms and decision metrics. Based on the information collected in the measurement step and algorithm designed, a certain path/gateway is decided in this step. Execution, the last step, illustrates which terminal devices make the decisions. The decisions can be made in nodes locally, or they can be made centrally and distributed to nodes.

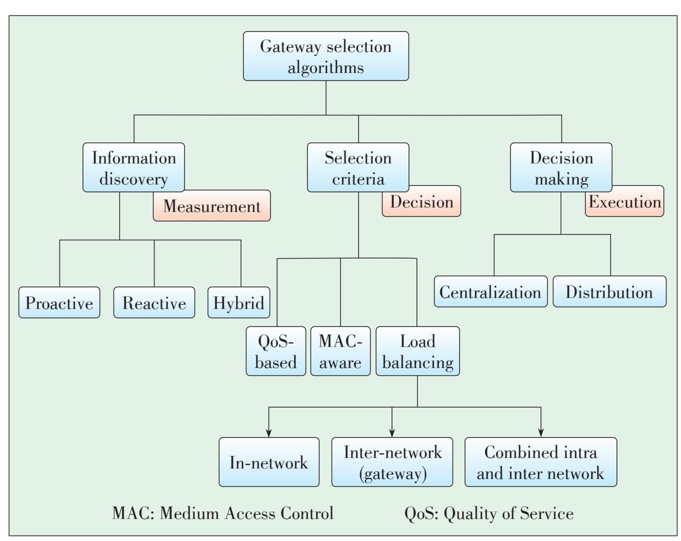
Fig. 1 illustrates the topology under consideration. The considered topology is formed from a MANET with large number of mobile nodes that can communicate with each other within communication range. Any mobile node can be the source node which requires external communication. Multiple gateways are available to provide external links and multiple paths are possible to route to gateways. In Fig. 1, for the source node s, there are currently three gateways available. They have different properties: gateway 1 and gateway 3 are able to handle more traffic and have higher capacity compared with gateway 2. And clearly, the node density around gateway 3 is higher than that around gateway 1 and gateway 2, which means it is highly possible that interference needs to be considered when it comes to gateway 3. In terms of location position, gateway 1 is the most far away one, which means hop distance delay is the largest theoretically. For gateway 2, there are two paths established with different hop distance and node density (transmission interference). Should the decisions made in node s or in other designed entity? How to make sure the validation and accuracy of these information? With all these information, which gateway and path should be selected to the node s to achieve the best network performance? Gateway selection aims at solving these questions.



▲Figure 1. Target MANET topology with multiple gateways and available paths.

3 State-of-the-Art Gateway Selection in an Integrated MANET

Various approaches have been proposed and in this section we present a literature review of the studies in gateway selection problem. Different gateway selection mechanisms are designed to achieve various functions and objectives in different systems. The studies are categorized in terms of various aspects as shown in **Fig. 2**. A basic category criterion is the way a gateway is discovered and how the gateway/path information is collected. For this purpose, proactive, reactive and hybrid mechanisms are proposed. Each mechanism has its own advantages. Another criterion is the means of decision-making for gateway selection in network level. The decision can be made



▲Figure 2. Category illustration of gateway selection algorithms.

in a central point (central controller) and a centralized mechanism is needed in this case. On the contrary, the decision can be made at local nodes that require the mechanisms to be distributed. The decision making we discuss here mainly focus on the terminal devices that make the decisions and the means of execution. Hence, the decision making is related to the execution process. Based on the objective of algorithm design, the studies of gateway selection can be categorized into QoS-aware, Medium Access Control (MAC)-aware and load balancing based group. **Table 1** summarizes the studies briefly in terms of different criteria. In the following part, these studies will be illustrated with more design details.

3.1 Decision Making

Based on the way of decision-making, the studies can be grouped as centralized solutions and distributed solutions. The most common methodology proposed to address gateway selection problem is to have a centralized controller that collects network information and runs an algorithm to determine gateways for mobile nodes which are grouped as centralized solutions. Gateways can also be responsible for making selection decisions, and we group these as gateway centralized studies. On the contrary, decision-making can occur in local nodes and based on this, the distributed solutions are designed for node-centric behavior. Based on the collected gateway and path information, mobile nodes make their own decisions for choosing which gateway to deliver traffic flows to.

The work in [4] considers 1) eliminating the gateway flapping, 2) avoiding harmful effect caused by gateway selection interference, and 3) balancing traffic at TCP flow level and improving the performance and fairness of flows. It is assumed that there is routing protocol executed to establish routes and perform routing within network. The network topology graph and routes are assumed to be obtained from the routing protocol. Gateways are connected to a central controller by wired network. When a set of flows passes through the gateway, the topology and path information is passed to the controller. Based on this, the controller executes a fast gateway selection algorithm and informs the gateway new association.

▼Table 1. Studies of gateway selection

Information discovery	Ref.	Selection criteria	Decision making
Proactive	[7]	Expected throughput; MAC link interference	Distributed
	[18]	Expected link quality; Interference ratio; Gateway load	
	[15]	Gateway capacity; Hop distance	Centralized
	[8]	Contention level; Congestion level; Hop distance	Gateway centralized
	[9]	Hop distance; Traffic volume	
	[14]	Gateway traffic load; Hop distance threshold	
Reactive	[4]	No. of Register nodes; Expected interference; Hop distance	Centralized
	[10]	Traffic volume	Distributed
	[19]	Gateway traffic load; Hop distance	
Hybrid	[5]	Path available period; Path capacity; Path latency	Distributed

The selection algorithm is periodically executed to refresh the associations and all the flows of a node have the same gateway association. The node-gateway association is decided according to two ranked lists. The unlocked nodes (nodes with no gateway association) are ordered according to the number of valid paths (gateways). The valid paths are ordered in ascending order based on the gateway load (number of nodes the gateway serves) and path cost (expected interference and hop distance). Unlocked nodes are assigned one by one to the gateway with minimum load and cost. Even though the aim is balancing the load at the flow level, this paper only explains node-gateway association.

Based on the network-wide information, centralized solutions are likely to determine the most suitable gateway for a certain node and make optimized decisions for all the considering nodes. However, they suffer from the fact that all the information has to be collected at a single point. In the topology-dynamically-changing network, to capture the latest overall network condition is challenging.

3.2 Selection Criteria

As mentioned, there are questions in gateway selection procedure: 1) which gateway to select for each traffic flow, 2) how to route data flows through multi-hop network paths established between selected gateways and mobile nodes. Various approaches to addressing these issues significantly affect the performance of the entire system. They can be designed to increase network throughput, improve QoS, and increase the efficiency of network resource usage. Considering the objective and emphasis of selection criteria, the studies can be classified into different categories: QoS-based, MAC-aware and Load balancing based algorithms. Selection criteria, which determine decision metrics, should be categorized to the decision process.

3.2.1 QoS Based Studies

To improve QoS, some performance metrics are designed to evaluate the quality of path/gateway based on network and gateway parameters. QoS based gateway selection mechanisms favours some network parameters, such as hop distance, end-to-end delay, link connectivity, and residual load capacity of gateways. Some early works were designed based on a single metric, however, single metric based algorithm sometimes needs to sacrifice another network performance. Hence, trade-off between network parameters should be taken into account and algorithms with multiple metrics are proposed.

The authors in [5] consider multiple QoS path parameters such as path availability period, available capacity and latency. In order to achieve considerations of both in-network and gateways, the parameter of the end-to-end path between a mobile node and a gateway node are computed. The path availability period is estimated based on mobility model and epoch length caused by its movement is considered. To avoid in-network bottlenecks, the residual load capacity of a path is formulated as the minimum available load capacity at any node along the path. Each node makes an independent decision on which gateway to choose based on the metric value to each gateway. The metric is based on a linear combination of the three terms.

The authors in [6] aim at balancing the inter/intra-MANET traffic load over multiple IGWs (Internet gateways). The paper provides a metric consisting three components: shortest Euclidean distance (hop count), inter-MANET traffic load and intra-MANET traffic load. The metric is a linear combination of these three components. The inter-MANET traffic load is represented by the number of registered MANET nodes (both local and visiting) at the IGW. The intra-MANET traffic load within the network is related to the optimal node density to delivery traffic successfully. However, the definitions of inter and intra-MANET traffic given in paper are partly redundant and not validated well.

3.2.2 MAC Based Studies

The selection decision over different routing paths depends on link performance indicators, which are influenced by the MAC parameters. Hence, the appropriate routing metric should adapt to the corresponding MAC parameters for efficient joint functioning. In [7], authors develop load balancing algorithm on top of expected throughput (ETP) routing metric. ETP is a MAC-aware routing metric which takes into account the capacity reduction of a link due to its interaction with other links in its contention domain. ETP takes into account all the existing flows on the paths at a node. It aims to determine a delay-optimal routing forest. Unlike the algorithms mentioned previously, this study does not explicitly use gateway advertisements or requests to gateways from nodes. All the information is achieved by neighboring messaged exchange. Control message overhead is sufficiently decreased. In this paper, the bit rate and the packet success probability of a link to all its potential parent nodes is available at current node. Based on these, next hop with minimum expected delay experienced by a bit is chosen which achieves local optimization. Benefiting from the distributed manner, it is expected to minimize the expected per bit delay over the entire network.

3.2.3 Load Balancing Based Studies

Due to the distributed behavior of mobile nodes, nodes in certain locations could be more vulnerable to congestion because they may have a higher number of neighbors or they are responsible for transmitting the majority of the packets. Gateway nodes can also be vulnerable to congestion because they aggregate all the network traffic. The congestion is caused by unevenly distribution of the traffic. Load balancing schemes are essential to distribute the traffic evenly and alleviate congestion. Furthermore, the load balancing schemes can be used to decrease the potential delay caused by congestions and improve the resource utilization. Therefore, load balancing is an important consideration for improving network performance. There are a number of studies that have been performed to focus on balancing network traffic and they differ in the metric used for representing path/gateway traffic load. These studies can be sub-categorized into: in-network load-balancing, gateway load-balancing and combine network and gateway load-balancing.

The in-network load-balancing mechanisms emphasize that the traffic between source nodes and gateways is distributed between a set of alternative paths in order to avoid congestions and improve network performance. Most of them focus on design of path performance metric. In [8], the contention level, congestion level, and hop distance are combined as the a selection metric to avoid areas which have high data traffic or many nodes contending for channel access. The authors in [9] propose a potential-field-based routing scheme to reflect both distance and traffic volume at each node to multiple gateways. Based on different path quality metrics, these studies distribute data traffic across the network. However, compared to mobile nodes, the gateway nodes are more vulnerable to congestions and easier to become bottlenecks because they aggregate all the network traffic.

The gateway load-balancing mechanisms aim at distributing traffic load among multiple gateways in order to reduce gateway load imbalance and to maximize the total network throughput. The related approaches mostly consider gateway load for selection and the gateway load is evaluated with different gateway parameters, such as actual amount of traffic [10], residual capacity [11], the number of active flows, average queue length [12], the number of registered nodes [13], and so on. The authors in [14] consider the gateway load and channel contention in cost function for gateway selection. Each node calculates its own cost contribution to a gateway by multiplying the hop distance by the reciprocal of the number of nodes whose traffic is forwarded. Each gateway receives route requests from all mobile nodes, calculates total costs (summation of the reciprocals of the number of nodes for which traffic is forwarded) and then includes the related information in its periodic advertisement message to mobile nodes. There is also a hop count threshold introduced, and the gateways that are more than a certain hop distance are not considered as a candidate. However, the cost function is not explained very well and is not justified as a good metric.

In [15], the authors design a practical, robust framework for load-balanced routing and gateway selection under variable traffic loads. In this architecture, end-to-end QoS provisioning and network resource management are deployed on a centralized controller called the Network Manager (NM). The NM is responsible for collecting information on the network topology and the available capacities of all links, maintaining a list of active flows. Two schemes are proposed: (i) to assign a given flow to the gateway with maximum residual capacity (MRC), (ii) to assign the flow to the gateway with maximum normalized capacity (MNC), which is the ratio of residual gateway capacity to hop distance. To always select gateways with a lighter load, these solutions may evenly distribute data packets among multiple gateways in average during a period. However, they do not take into account the effects of route-switching, which may cause gateway flapping and also degrade in-network performance [16].

There are only a few studies about combining load balancing within the network and among gateways. In [17], gateway load, route interference and expected link quality are combined as metric for nodes to choose the best gateway. The gateway load is defined as the average queue length at the network interface of accessing gateway. To estimate the gateway load accurately, an exponential weighted moving average is used to smooth out variations. Route interference is defined as the sum of the link interferences along the route. Link interference is the maximum of the node interference, where node interference is defined as the percentage of the time nodes sense wireless activities in the channel. Finally, the expected link quality is defined as the forward link packet delivery ratio. Using the combined metric (of gateway load, route interference, and expected link quality), the aim is to evenly distribute traffic across the network and equalize the traffic load on gateways. However, in real scenarios, gateways may have different capacities. Even though the data traffic is uniformly distributed between multiple gateways, a few gateways may still be overloaded. Hence, the utilization, which is defined as ratio of traffic load over maximum capacity of a gateway, should be considered in future work. Moreover, this study uses proactive approach for information discovery, where gateways periodically broadcast advertisement packets including gateway load and mobile nodes periodically broadcast probe packets including node interference factor and packet delivery ratio. A proactive approach may keep routes up to date, but it increases interference due to large number of routing packets.

In [18], a source based gateway selection scheme is proposed and it combines path metric and gateway load. Path metric is a function of link metric, which is defined as a combination of expected link quality and interference ratio. Expected link quality is the success rate of transmitted probe packets. The interference ratio is the ratio of the sum of interference power from all interfering nodes over the maximum tolerable interference at the receiver. The gateway load is based on the average interface queue length and represented as gateway capacity. The gateway capacity is the sum of the capacities of three interfaces in each gateway, where the capacity of each interface is defined as subtraction of maximum interface queue capacity and current interface queue length. This study also proposes a waiting time before switching to a better path because an immediate switch may cause congestion and oscillations when several source nodes switch to a better path at the same time. However, it does not consider that the better path may change during this waiting time period. Even though the gateway load is part of selection metric, this study focuses on decreasing packet loss/retransmission, instead of balancing network load. Another issue of this study is that only source nodes are designed and implemented with gateway/path selection algorithm. In practical scenarios, source nodes are normally far away from gateways and they may not be sensitive enough to catch the latest network conditions which dynamically change. Hence, the intermediate nodes should also be able to make decisions based on the latest path information to improve selection accuracy.

3.3 Gateway and Information Discovery

The gateway/path discovery is about information collection and the accuracy of path parameters is an important design objective. The functionality mechanisms (QoS aware, MAC-aware and load balancing) are usually built on top of discovery algorithm. Things that need to be considered with the different methods for gateway/path discovery are network control overhead packets, computation complication of algorithm, and initial delay of path setup.

In [6] and [14], gateways proactively broadcast their current information of the number of registered nodes and network topology periodically. For proactive behavior, there will be large control packets transmitting in network, which consumes network resources. Besides, the path/gateway information may not be updated frequently enough to keep up to date. The advantage is the routing information will be always available in mobile nodes and packets can be redirected immediately when it arrives at the node.

The authors in [19] use reactive behavior to collect the path/gateway information, which means only when the route to a certain destination is required, the mobile node will initiate a route request. The current path/gateway condition will be integrated into the route reply packets. Control packets in this case include route request and route reply packets which will be much smaller than those in the proactive algorithm. On the other hand, the route initiation takes time, which increases the end-to-end delay of packet transmission.

In [5], the authors use a hybrid gateway discovery algorithm. The proactive and reactive zones are divided by the number of hops from mobile nodes to gateway node. Each gateway periodically broadcast the gateway advertisement message which contains its parameters within a proactive region of k hops. The nodes in reactive zone discover the gateway node and path information by sending gateway discovery message. In this case, the number of control messages will be reduced compared with the proactive scheme. Also, the nodes within k-hop distance have less route establishment time compared to the reactive scheme. However, it requires more complicated computation work in this scheme and how to define the optimal proactive region is critical in this scheme.

4 Open Issues and Challenges for Future Studies

Integration of MANET and other networks has been well studied, but when it comes to gateway selection, there are still some additional issues that must be taken into account and addressed. In this section, the challenges and issues are explained and discussed. **Table 2** summaries the challenges and future directions.

▼Table 2. Challenges and directions for gateway selection

Challenge	Problem Statement	Direction
Route flapping	Synchronization rerouting; Traffic unevenly distribution	To evaluate route change effects; Traffic control
Performance trade-off	Multiple objective; Resource and cost conflict	To evaluate costs (individual and combine objectives)
Path information update	High accuracy cost, energy and network resource consuming	To evaluate information update cost; Maximize frequency and limit resource consuming

4.1 Routing Flapping

The most distinguished feature of ad hoc networks is that the self-organization which makes it impossible to achieve accurate, instant-react routing management in-network. Most of the protocols are designed in a distributed manner, and there will be latency for nodes' reaction to changes in the network topology. It is likely that at one time all the nodes find out the current default gateway is overloaded, so they turn to register in another gateway. This leads to overloading on the targeted gateway. The overloading cannot be alleviated until the time that the overloaded information of this gateway is known by the majority of mobile nodes. At this time, the other gateway starts to overload again. This synchronized rerouting problem causes routing flapping (also referred to as the ping pong effect). It is hard to maintain an absolutely balance in load distribution, however, optimal schemes can be applied to alleviate this flapping effect. There must be coordination between gateways and mobile nodes to intelligently select the gateway and alleviate flapping.

Furthermore, a route change in a node may affect others, leading to affect overall network performance. The more aggressive of the gateway switching/route change is, the more likely to be poor network performance is. The trade-off between route decision and network performance also need to be considered. Thus, how to evaluate the effect of route changes and how to control the changes to achieve the best performance gain must be further studied.

4.2 Performance Trade-Off

With respect to the network view, there is a cost if additional improvements are made beyond the natural network behavior. To achieve realistic situation, multiple-objectives are normally required, for example, to achieve load-balancing and multi-QoS at the same time. However, when a certain objective is achieved, normally some other may not be accomplished. In this case, how to coordinate the behavior and performance of different metrics is crucial and the trade-off of different achievements needs to be considered.

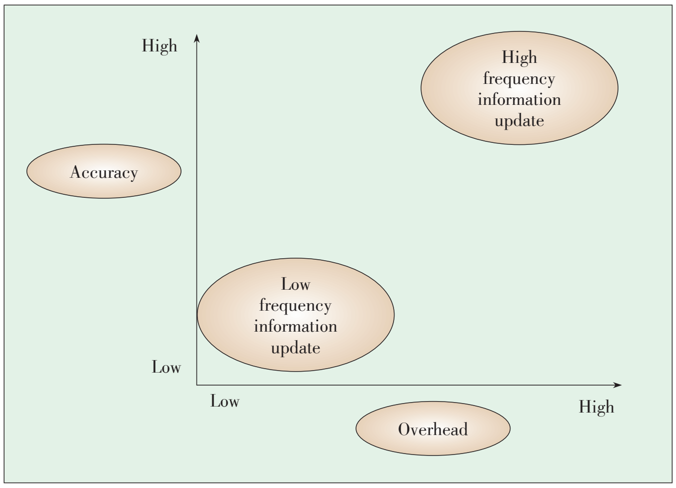
For example, when path reliability is of high priority, a more reliable, longer distance path may be chosen for higher reliability. In this case, the overall data delivery ratio will be increased but also the end-to-end delay. Due to large amount of data aggregated, it is likely to be congested when the nodes have a large number of neighbor nodes or the nodes are on the shortest path. To alleviate the congestion, a less traffic path with longer hop distance can be selected. For example, gateway 1 will be chosen instead of gateway 2 and gateway 3 in Fig. 1. Load-balancing is achieved but delay is increased as the cost.

Different QoS and priorities need to be considered for various applications and scenarios/topologies. In terms of applications, the voice message values the transmission delay and jitter. On the other hand, file transmission prefers high delivery ratio other than delay. When it comes to application scenarios, emergency response integrated system requires both delay and delivery ratio to be high level performance. Most of the mechanisms are designed to be effective in a specific scenario. The challenging issue in this case is the design of coefficients that should be able to not only achieve promising performance in theory but affect the realistic situation. Besides, the goal is to make best use of network resource and sacrifice the least.

Furthermore, the most challenging part is how to design an adaptive mechanism that is able to cope with different cases through its own learning/adaptive function and achieve the related goals.

4.3 Information Update Frequency (Trade-off Between Accuracy and Cost)

The route/gateway selection decision depends on current path conditions. Parameters of path conditions are collected from nodes by different information discovery methods. The frequency of path parameter updates determines how accurately a node can adapt to changes. There is a trade-off between acquired accuracy and resource efficiency including overhead packets and energy consumption (**Fig. 3**). More frequent updates may achieve more accuracy when adapting to network dynamics, but it may also cause higher control messages consume and higher energy consumption. If the update is not frequent enough, the changes of state of gateways/paths may not



▲Figure 3. Trade-off between control overhead and information accuracy.

be informed to mobile nodes in time. The information stored in nodes could be out of date. Hence, how to limit the update frequency and at the same time achieve the desirable accuracy needs to be investigated. In addition, route changes may affect the diffusion of path information. Hence, to make sure the information is disseminated efficiently, the route change effect should also be considered as one aspect.

In multihop Wireless Sensor Networks (WSNs) with static gateways, the nodes close to the gateways are more likely to deplete their battery supplies before the far-away nodes due to the intersection of multihop routes and concentration of data traffic towards the gateways. To achieve uniformity of energy consumption, the usage of mobile gateways is proposed and explored recently [20], [21]. As the gateway moves, the hotspot nodes around the gateway change and energy increaseingly drains through the network. Furthermore, to harvest multiple-input multiple-output (MIMO) gains in WSNs, several nodes which are equipped with one or more antennas emulate a multi-antenna node and independent paths between the transmit and receive sides can be realized by having multiple spatially separated nodes [22], [23].

5 Conclusions

To provide inter-connection between intrinsically different networks, data aggregation function and network performance guarantee, the gateway in an integrated system is vital to satisfy requirements. The gateway selection problem can be addressed as a centralized optimization problem, where performance metrics (e.g. end-to-end delay, overall network throughput, etc.) should be maximized. To achieve the objectives, the network topology and traffic flows need be assumed to be static and known. However, either network topology or traffic flows are uncertain and highly dynamic in the system. The gateway selection problem, on the other hand, can be addressed as a network performance improvement problem in terms of different routing metrics (e.g., traffic load, path delay, multiple QoS metrics, etc.). The information is gathered in a distributed learning way that may decreases timeliness. Besides, in order to achieve local optimization, certain paths may be sabotaged by related nodes.

Algorithms for future designs should be elaborated to cope with realistic scenarios and deal with the challenges stated in the last section. Update of path/gateway quality parameters must be designed carefully and efficiently, in which case it guarantees the information accuracy and achieves performance gain with the least network cost. Due to dynamic topology and real-time traffic condition, the route decisions should be dynamic and based on current information. It is desirable to design the algorithm manageable and adjustable to different network conditions. Furthermore, the problem can be further extended to gateway selection in multi-domain wireless networks [24], [25], where the inter-domain costs need to be considered.

References

[1] T. Taleb, Y. H. Aoul, and T. Ahmed, "Challenges, opportunities, and solutions for converged satellite and terrestrial networks," *IEEE Wireless Communications*, vol. 18, no. 1, pp. 46–52, Feb. 2011. doi: 10.1109/MWC.2011.5714025.
[2] A. Aijaz, H. Aghvami, and M. Amani, "A survey on mobile data offloading: technical and business perspectives," *IEEE Wireless Communications*, vol. 20, no. 2, pp. 104–112, Apr. 2013. doi: 10.1109/MWC.2013.6507401.
[3] A. Oliveira, Z. Sun, M. Monier, *et al.*, "On optimizing hybrid ad-hoc and satellite networks—the MONET approach," in *Future Network and Mobile Summit*, Florence, Italy, 2010, pp. 1–8.
[4] J. J. Galvez, P. M. Ruiz, and A. F. G. Skarmeta, "Responsive on-line gateway load-balancing for wireless mesh networks," *Ad Hoc Networks*, vol. 10, no. 1, pp. 46–61, Jan. 2012.
[5] S. H. Bouk, I. Sasase, S. H. Ahmed, and N. Javaid, "Gateway discovery algorithm based on multiple qos path parameters between mobile node and gateway node," *Journal of Communications and Networks*, vol. 14, no. 4, pp. 434–442, 2012.
[6] Q. Le-Trung, P. E. Engelstad, T. Skeie, and A. Taherkordi, "Load-balance of intra/inter-MANET traffic over multiple internet gateways," in *MoMM' 08*, Linz, Austria, 2008, pp. 50–57. doi: 10.1145/1497185.1497200.
[7] V. Mhatre, H. Lundgren, F. Baccelli, and C. Diot, "Joint MAC-aware routing and load balancing in mesh networks," *ACM CoNEXT' 07*, New York, USA, 2007, article no. 19. doi: 10.1145/1364654.1364679.
[8] Y. Fu, K.-M. Chan, K. S. Tan, and B. S. Yeo, "Multi-metric gateway discovery for MANET," in *IEEE 63rd Vehicular Technology Conference*, Melbourne, Austrialia, 2006, pp. 777–781. doi: 10.1109/VETECF.2006.1682930.
[9] S. Jung, M. Kserawi, D. Lee, and J.-K. K. Rhee, "Distributed potential field based routing and autonomous load balancing for wireless mesh networks," *IEEE Communications Letters*, vol. 13, no. 6, pp. 429–431, Jun. 2009. doi: 10.1109/LCOMM.2009.090110.
[10] Y. Miao, Z. Sun, F. Yao, *et al.*, "Study on research challenges and optimization for internetworking of hybrid manet and satellite networks," in *PSATS*, Toulouse, France, 2013, pp. 90–101. doi: 10.1007/978-3-319-02762-3_8.
[11] C.-F. Huang, H.-W. Lee, and Y.-C. Tseng, "A two-tier heterogeneous mobile ad hoc network architecture and its load-balance routing problem," *Mobile Networks and Applications*, vol. 9, no. 4, pp. 379–391, Aug. 2004. doi: 10.1023/B:MONE.0000031605.84447.1d.
[12] D. S. Nandiraju, L. Santhanam, N. Nandiraju, and D. P. Agrawal, "Achieving load balancing in wireless mesh networks through multiple gateways," in *IEEE MASS*, Vancouver, Canada, 2006, pp. 807–812. doi: 10.1109/MOBHOC. 2006. 278655.
[13] J. Shin, H. Lee, J. Na, *et al.*, "Load balancing among internet gateways in ad hoc networks," in *IEEE 62nd Vehicular Technology Conference*, Dallas, USA, 2005, pp. 1677–1680. doi: 10.1109/VETECF.2005.1558227.
[14] K. Tada and M. Yamamoto, "Load-balancing gateway selection method in multi-hop wireless networks," in *IEEE GLOBECOM*, Honolulu, USA, 2009, pp. 1–6. doi: 10.1109/GLOCOM.2009.5425325.

[15] E. Ancillotti, R. Bruno, and M. Conti, "Load-balanced routing and gateway selection in wireless mesh networks: design, implementation and experimentation," in *IEEE WoWMoM*, Montreal, Canada, 2010, pp. 1-7. doi: 10.1109/WOWMOM.2010.5534991.
[16] V. Pham, E. Larsen, P. E. Engelstad, and Ø. Kure, "Performance analysis of gateway load balancing in ad hoc networks with random topologies," in *MobiWAC' 09*, Tenerife, Spain, Oct. 2009, pp. 66-74. doi: 10.1145/1641776.1641787.
[17] U. Ashraf, S. Abdellatif, and G. Juanole, "Gateway selection in backbone wireless mesh networks," in *IEEE WCNC*, Budapest, Hungary, 2009, pp. 2548–2553. doi: 10.1109/WCNC.2009.4917735.
[18] M. Boushaba, A. Hafid, and M. Gendreau, "Source-based routing in wireless mesh networks," *IEEE Systems Journal*, vol. PP, no. 99, pp. 1–9, 2014. doi: 10.1109/JSYST.2014.2317453.
[19] Y. Chaba, R. B. Patel, and R. Gargi, "Efficient multiple DYMO routing protocol with gateway selection for hybrid manets," *International Journal of Computer Theory and Engineering*, vol. 4, no. 1, Aug. 2012. doi: 10.7763/IJCTE.2012.V4.541.
[20] S. Yu, B. Zhang, C. Li, and H. Mouftah, "Routing protocols for wireless sensor networks with mobile sinks: a survey," *IEEE Communications Magazine*, vol. 52, no. 7, pp. 150–157, Jul. 2014. doi: 10.1109/MCOM.2014.6852097.
[21] C. Tunca, S. Isik, M. Donmez, and C. Ersoy, "Distributed mobile sink routing for wireless sensor networks: A survey," *IEEE Communications Surveys Tutorials*, vol. 16, no. 2, pp. 877–897, 2014. doi: 10.1109/SURV.2013.10013.00293.
[22] D. N. Nguyen and M. Krunz, "A cooperative MIMO framework for wireless sensor networks," *ACM Transactions on Sensor Networks*, vol. 10, no. 3, May 2014.
[23] Q. Gao, Y. Zuo, J. Zhang, and X.-H. Peng, "Improving energy effciency in a wireless sensor network by combining cooperative MIMO with data aggregation," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 8, pp. 3956–3965, Oct. 2010. doi: 10.1109/TVT.2010.2063719.
[24] Y. Song, S. H. Y. Wong, and K.-W. Lee, "Optimal gateway selection in multi-domain wireless networks: a potential game perspective," in *17th Annual International Conference on Mobile Computing and Networking*, Las Vegas, USA, Sept. 2011, pp. 325–336. doi: 10.1145/2030613.2030650.
[25] S. Zhong and Y. Zhang, "How to select optimal gateway in multidomain wireless networks: alternative solutions without learning," *IEEE Transactions on Wireless Communications*, vol. 12, no. 11, pp. 5620-5630, Nov. 2013. doi: 10.1109/TWC.2013.100113.121828.

Manuscript received: 2015-09-02

Biographies

Ye Miao (y.miao@surrey.ac.uk) received her BSc in Electronic Information Science and Technology from China Agricultural University, China, in 2010 and MSc in Mobile and Satellite Communication from University of Surrey, UK, in 2011. She is currently a PhD student at the 5G Innovation Centre, Institute for Communication Systems (ICS), University of Surrey, UK. Her research interests are QoS and routing solutions in MANETs, integrated MANET-Satellite networks and WSNs.

Zhili Sun (z.sun@surrey.ac.uk) is the chair of Communication Networking and a Professor at the 5G Innovation Centre, ICS, University of Surrey, UK. He has been with University of Surrey since 1993. He got his PhD degree in Computer Science from Lancaster University, UK, in 1991. He worked as a postdoctoral research fellow with Queen Mary University of London from 1989 to 1993. He has been principle investigator and technical co-coordinator in many projects within the EU framework programs, ESA, EPSRC, and industries, and published over 125 papers in international journals, book chapters, and conferences. He is the sole author of the book titled *Satellite Networking—Principles and Protocols* by Wiley in 2005, a contributing editor of the book *IP Networking Over Next Generation Satellite Systems* published by Springer in 2008, and a contributing editor of the 5th edition of the text book *Satellite Communications Systems—Systems, Techniques and Technology* published by Wiley in 2009. His research interests include wireless and sensor networks, satellite communications, mobile operating systems, traffic engineering, Internet protocols and architecture, QoS, multicast and security.

Ning Wang (n.wang@surrey.ac.uk) received his B.Eng. (honours) degree from Changchun University of Science and Technology, China, in 1996, his M.Eng. degree from Nanyang University, Singapore, in 2000, and his PhD degree from University of Surrey, UK, in 2004. He is a senior lecturer at the 5G Innovation Centre, ICS, University of Surrey, UK. His research interests mainly include information centric networking (ICN), content and data caching management, network optimisation techniques and QoS.

DOI: 10.3969/j. issn. 1673-5188. 2015. 04. 008
http://www.cnki.net/kcms/detail/34.1294.TN.20150914.1512.002.html, published online September 14, 2015

Screen Content Coding with Primary and Secondary Reference Buffers for String Matching and Copying

Tao Lin, Kailun Zhou, and Liping Zhao
(Tongji University, Shanghai 200092, China)

Abstract

A screen content coding (SCC) algorithm that uses a primary reference buffer (PRB) and a secondary reference buffer (SRB) for string matching and string copying is proposed. PRB is typically the traditional reconstructed picture buffer which provides reference string pixels for the current pixels being coded. SRB stores a few of recently and frequently referenced pixels for repetitive reference by the current pixels being coded. In the encoder, searching of optimal reference string is performed in both PRB and SRB, and either a PRB or SRB string is selected as an optimal reference string on a string-by-string basis. Compared with HM-16.4+SCM-40 reference software, the proposed SCC algorithm can improve coding performance measured by bit-distortion rate reduction of average 4.19% in all-intra configuration for text and graphics with motion category of test sequences defined by JCT-VC common test condition.

Keywords

HEVC; Image Coding; Screen Content Coding; String Matching; Video Coding

1 Introduction

With continuous and rapid advancements in communications, networking, computers, semiconductors, and displays technologies, screen content coding (SCC) becomes a key technology for some emerging popular applications such as cloud computing, cloud-mobile computing, Wi-Fi display, smartphone and tablet second display, ultra-thin clients, remote desktops, and screen sharing [1]–[3]. The challenging requirement of SCC is to achieve both ultra-high visually lossless quality and ultra-high compression ratio up to 300:1–3000:1. In recent years, SCC has attracted increasing attention of researchers from both academia and industry [4]–[30]. ISO/IEC MPEG and ITU-T VCEG have set up the Joint Collaborative Team (JCT) for a joint project of developing a SCC version of High Efficiency Video Coding (HEVC, also known as H.265) standard and a joint call-for-proposal has been issued in January 2014 [14].

Characteristics of computer screen pictures are quite different from those of traditional pictures, for which traditional block-matching and transform-based hybrid coding technique can compress efficiently. In general, computer screen pictures are essentially a mix of two-type contents: discontinuous-tone content such as text, chart, graphics, and icon, and continuous-tone content such as photograph, movie/TV clips, natural picture video sequences, and sophisticated light-shaded and texture-mapped virtual-reality scenes. Continuous-tone content features relatively smooth edges, complicated textures, and thick lines with virtually unlimited colors. In contrast, discontinuous-tone content features very sharp edges, uncomplicated shapes, and thin lines with few colors, even one-pixel-wide single-color lines. General screen content pictures are often seen in web browsing, video conferencing with document sharing, office document editing, engineering drawing, hardware design engineering, software programming, gaming, map navigating, address direction searching, and other computer use cases.

Actually, typical computer screens for daily use are often rich in small and sharp bitmap structures such as text, menu, icon, button, slide-bar, and grid. There are usually many similar or identical patterns in a screen picture. These similar or identical patterns may have very different sizes from a few pixels to a few thousands of pixels and very different shapes. Existing techniques such as intra-prediction and intra block copy (IBC) [15]–[17] is not always efficient to code similar or identical pattern within a picture, because they use only 1-D pattern or 2-D pattern of a few fixed sizes and a few fixed rectangle or square shapes.

In order to address the issue and improve the coding efficiency of repeated patterns with different sizes and shapes, this paper proposes a coding technique based on string-matching (also called intra string copy or ISC). It provides a very generic and flexible solution to string matching of variable sizes and different shapes.

Two reference buffers are used in the proposed technique. One is primary reference buffer (PRB) that is typically the traditional reconstructed picture buffer to provide reference string pixels for the current pixels being coded. The other is secondary reference buffer (SRB), a lookup table (LUT) storing a few of recently and frequently referenced pixels for repetitive reference by the current pixels being coded. For any starting pixel in an encoding process, searching of optimal reference string is performed in both PRB and SRB. Either a PRB string or a SRB

This work was supported in part by National Natural Science Foundation of China under Grant No. 61201226 and 61271096, Natural Science Foundation of Shanghai under Grant No. 12ZR1433800, and Specialized Research Fund for the Doctoral Program under Grant No. 20130072110054.

string is selected as an optimal reference string on a string-by-string basis. A PRB string has two string-matching parameters. One is an offset that specifies the relative position (2D coordinates), i.e. 2D displacement from the reference string to the current string. The other is a length that specifies the number of pixels in the reference string. It is obvious that the reference string and the current string have the same length. A SRB string also has two string-matching parameters: an index that specifies the address of the reference pixel in the SRB and a length that specifies the duplication count of the reference pixel to reconstruct the current string. The string-matching parameters are then entropy-coded and put into the final bit-stream. At the decoder side, the string-matching parameters are actually string-copying parameters. In the decoding process, the string-matching (i.e. string-copying) parameters are parsed, decoded, and used to obtain the reference string pixels from either PRB or SRB. The values of the reference string pixels are then assigned to the current pixels being decoded to reconstruct the current string.

In Section 2, a general architecture of string-matching enhanced coding system using both PRB and SRB is proposed and the details of horizontally and vertically scanned 2D-shape-preserved matching modes are described. Section 3 presents a flexible hash-table framework specially designed for speeding up reference string searching in PRB. Section 4 describes a method to seamlessly mix PRB search with SRB search inside a coding unit (CU), and to select one-by-one either an optimal PRB string or an optimal SRB string based on average rate-distortion (RD) cost evaluation. In section 5, the proposed technique is compared with HM-16.4+SCM-40 reference software [31] that is developed based on HEVC Software Model HM-16.4 by JCT for SCC testing. The experimental results show that the proposed technique can achieve significant bit-distortion rate (BD-rate) [32]–[33] reduction in lossy coding and bit-rate saving in lossless coding using the test suite defined by JCT Common Test Condition[30]. We conclude the paper and introduce the future work on string-matching coding technique in section 6. In this paper, the proposed SCC algorithm is described for pictures of YUV4:4:4 and RGB formats, but it can also be applied to pictures of YUV4:2:2 and YUV4:2:0 formats with some modifications. This paper is partially based on JCT-video coding (VC) documents [18]–[27].

2 String-Matching Enhanced Coding System

A general architecture and major components of string-matching enhanced coding system are shown in **Fig. 1**.

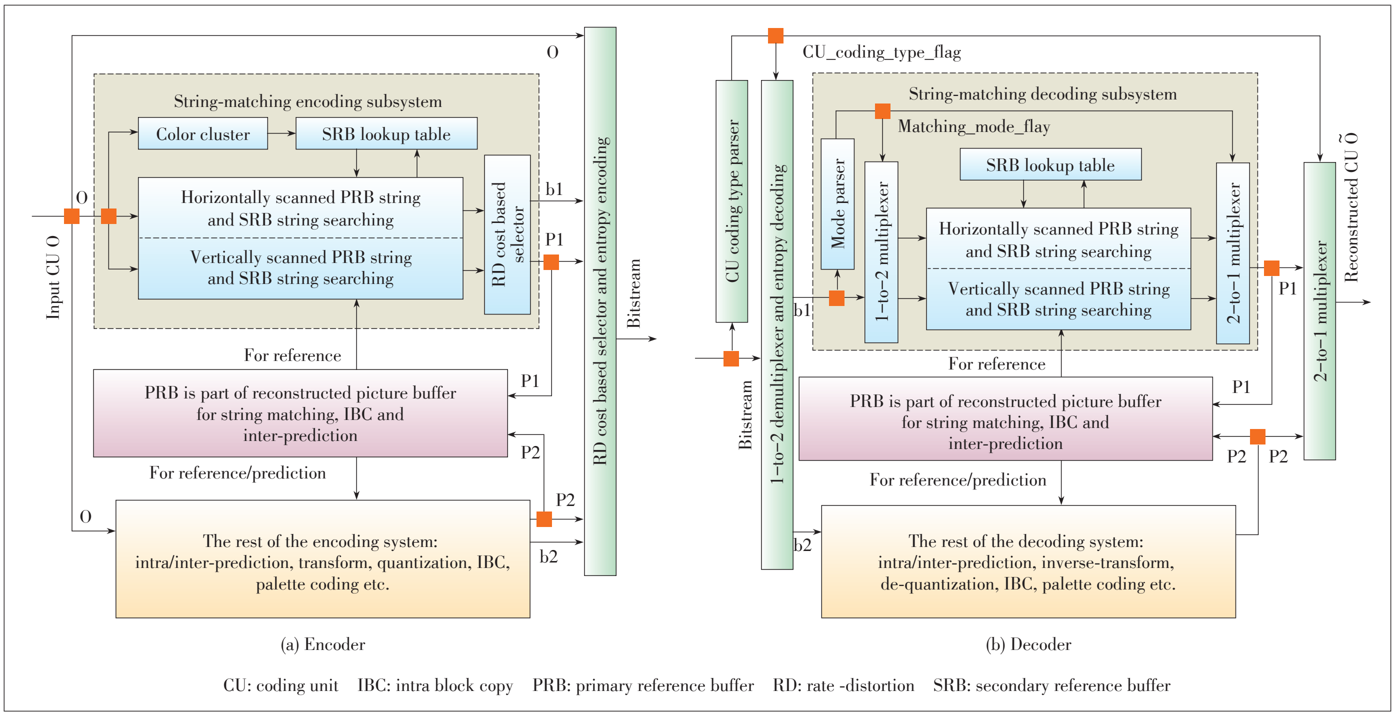
In the encoder of the string-matching enhanced coding system, the string-matching encoding subsystem performs color clustering, PRB string searching, and SRB string searching. The rest of the encoding system performs other traditional encoding operations such as intra-prediction, inter-prediction, transform, quantization, entropy encoding, IBC, and palette coding. The input CU O is fed to both string-matching encoding subsystem and the rest of the encoding system. The string-matching encoding subsystem codes O to generate coded bitstream b_1 and reconstructed CU P_1. The rest of the encoding system also codes O to generate coded bitstream b_2 and reconstructed CU P_2. O, P_1, b_1, P_2, and b_2 are sent to RD-cost-based selector that calculates the RD cost of two results and selects the one with the best RD performance as the final coding result for the CU. The corresponding coded bitstream b_1 or b_2 is selected and put into the output bitstream. Also, the corresponding reconstructed CU P_1 or P_2 is stored in reconstructed picture buffer, which is shared by both string-matching encoding subsystem as PRB for string-matching and the rest of the encoding system for inter prediction and IBC. The input CU O is also fed to a color cluster unit to obtain a few of most frequently occurred pixel colors. Some of the colors are put into the SRB LUT for SRB string searching.

In the decoder, the input bitstream is first parsed by CU coding type parser to get the CU_coding_type_flag that indicates whehter the current CU being decoded is coded by the proposed string-matching technique or by traditional coding techniques. If the CU is coded by the string-matching technique, the CU layer bitstream is sent to the string-matching decoding subsystem that decodes and reconstructs the CU P_1, which is stored in reconstructed picture buffer and is also the final reconstructed CU output $\tilde{O}$. If the CU is coded by traditional coding techniques, the CU layer bitstream is sent to the rest of the decoding system that performs traditional decoding tasks such as intra-prediction, inter-prediction, inverse-transform, de-quantization, IBC, and palette decoding. The CU P_2 is eventually reconstructed, which is stored in reconstructed picture buffer and is also the final reconstructed CU output $\tilde{O}$.

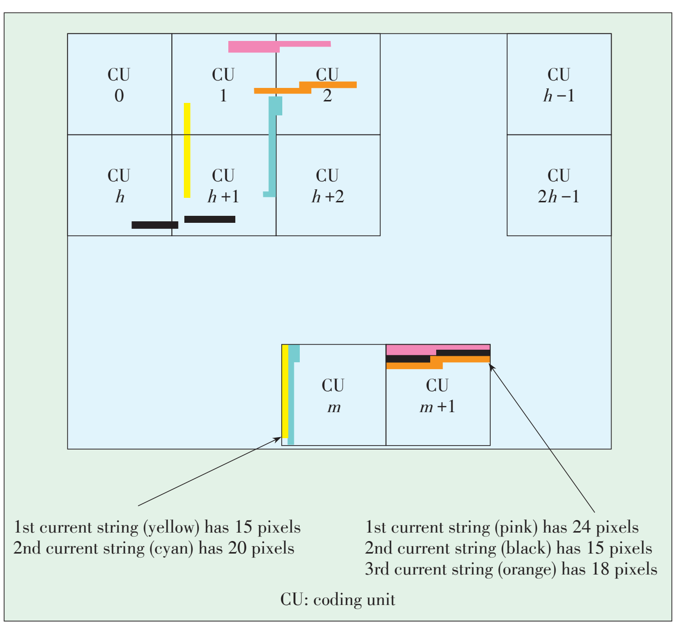
The string-matching coding subsystem has two CU level matching modes: horizontally scanned 2D-shape-preserved matching and vertically scanned 2D-shape-preserved matching (**Fig. 2**). A CU is pre-coded by both the modes. The mode minimizing RD cost is selected as the final string-matching mode for the CU. The CU size in Fig. 2 is 16x16 pixels in the packed pixel format (e.g. a Y sample is followed by a U sample and then a V sample, or a G sample is followed by a B sample and then an R sample).

The vertically scanned 2D-shape-preserved matching mode is used to code CU *m* in Fig. 2. In this mode, PRB is treated as a 2D plane while CU *m* is considered a vertically scanned 1D pixel string. Starting from the 1st pixel of CU *m*, the encoder searches the optimal (e.g. longest in lossless case) string in the PRB searching range that matches the pixel string in CU and also keeps exactly the same 2D shape as the pixel string in CU *m*. The string found in the searching range is called a reference string and the pixel string in CU *m* is called current string. Fig. 2 shows the first two current strings in CU *m* and their corresponding reference strings:

1) The 1st reference (and current) string in yellow has 15 pix-



▲**Figure 1. String-matching enhanced coding system architecture.**



▲**Figure 2. Two matching modes of string-matching coding.**

els and the 2D-shape-preserved reference string is located across CU 1 and CU h+1;

2) The 2nd reference (and current) string in cyan has 20 pixels and the 2D-shape-preserved reference string is located across CU 1, CU 2, and CU h+1.

The horizontally scanned 2D-shape-preserved matching mode is used to code CU m+1 in Fig. 2. In this mode, PRB is treated as a 2D plane while CU m+1 is considered to be a horizontally scanned 1D pixel string. Starting from the 1st pixel of CU m+1, the encoder searches the optimal string in the PRB searching range that matches the pixel string in CU m+1 and also keeps exactly the same 2D shape as the pixel string in CU m+1. The string found in the searching range is called a reference string and the pixel string in CU m+1 is called the current string. Fig. 2 shows the first three current strings in CU m+1 and their corresponding reference strings:

1) The 1st reference (and current) string in pink has 24 pixels and the 2D-shape-preserved reference string is located across CU 1 and CU 2;

2) The 2nd reference (and current) string in black has 15 pixels and the 2D-shape-preserved reference string is located across CU h and CU h+1;

3) The 3rd reference (and current) string in orange has 18 pixels and the 2D-shape-preserved reference string is located across CU 1 and CU 2.

There are usually two types of pixel scanning methods (and paths) available for a CU or a largest coding unit (LCU). One is raster-scan: a CU or LCU is scanned line by line, either horizontally or vertically, and all lines have the same scan direction (Fig. 2). The other is traverse-scan: a CU or LCU is also scanned line by line, either horizontally or vertically, but odd lines and even lines have opposite scan direction. Once a scanning method (and path) is determined, all pixels inside a CU or LCU are lined up following the scanning path. Thus, starting from a current pixel being coded inside a CU $P_{m,n}$, a current string cur$S_{m,n}$ can be defined following the scanning path. The pixels of cur$S_{m,n}$ are designated as $S_{m,n}(0)$, $S_{m,n}(1)$, ... , $S_{m,n}(L-1)$, following the order of the pixels on the scanning path, where L is the length of the string in unit of pixel. Using the designation, cur$S_{m,n}$ and its pixels can be expressed as cur$S_{m,n}$ = {$S_{m,n}(k)$: $0 \leqslant k < L$}. Given a current string cur$S_{m,n}$ and a reference pixel $P_{i,j}$, a reference string ref$S_{i,j}$ = {$S_{i,j}(k)$: $0 \leqslant k < L$} is the string that starts from $P_{i,j}$, i.e. $S_{i,j}(0) = P_{i,j}$, and has exactly the same 2D shape and scanning order as cur$S_{m,n}$.

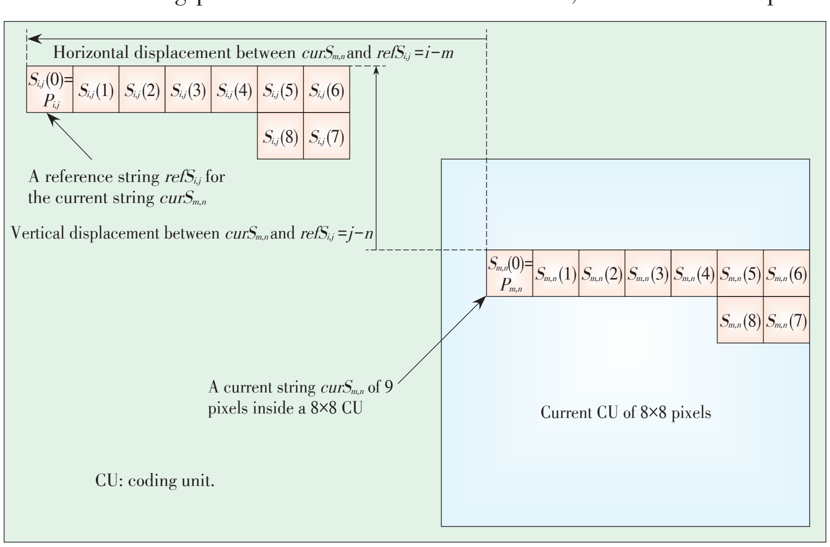
Fig. 3 illustrates an example of the current string cur$S_{m,n}$ = {$S_{m,n}(k)$: $0 \leqslant k < L$} starting from the current pixel $P_{m,n}$ inside a 8×8 CU of horizontal traverse-scan and its reference string ref$S_{i,j}$ = {$S_{i,j}(k)$: $0 \leqslant k < L$} starting from the reference pixel $P_{i,j}$. Both cur$S_{m,n}$ and ref$S_{i,j}$ have identical 2D shape and length of $L = 9$.

Obviously, given a current pixel $P_{m,n}$, the current string cur$S_{m,n}$ of length L can be uniquely specified within a CU with known size and scanning method, and a reference string ref$S_{i,j}$ is uniquely specified by a displacement vector ($\boldsymbol{DV}_h$, $\boldsymbol{DV}_v$) = (i-m, j-n). In Fig. 3, $\boldsymbol{DV}_h = i - m$ is the horizontal displacement between cur$S_{m,n}$ and ref$S_{i,j}$, and $\boldsymbol{DV}_v = j - n$ is the vertical displacement between cur$S_{m,n}$ and ref$S_{i,j}$.

A matching reference string ref$S_{i,j}$ = {$S_{i,j}(k)$: $0 \leqslant k < L$} for a given current string cur$S_{m,n}$ = {$S_{m,n}(k)$: $0 \leqslant k < L$} satisfies the constraint that the difference between $S_{i,j}(k)$ and $S_{m,n}(k)$ is within a predetermined threshold for all k. One common option used for difference measurement is the absolute difference $|S_{i,j}(k)_Y - S_{m,n}(k)_Y|$, $|S_{i,j}(k)_U - S_{m,n}(k)_U|$, and $|S_{i,j}(k)_V - S_{m,n}(k)_V|$, where the subscripts Y, U, V designate the Y color component, U color component, and V color component for the corresponding pixels $S_{i,j}(k)$ or $S_{m,n}(k)$.

3 String-Matching Oriented Hash-Table Framework

String-matching coding performance heavily depends on the PRB searching range. The larger the searching range is, the better the coding performance can achieve. However, the longer the searching time is, the more the computing power requires. Hash-table can be used to speed up reference string searching. Therefore, the key to design a practical string-matching encoding subsystem is to have a single and efficient string searching oriented hash-table that should work and speed up the searching in all two matching modes.



▲**Figure 3. An example of current string and its reference string.**

In the string-matching oriented hash-table framework, the basic role of a hash-table is to quickly find the first matching reference pixel in the PRB searching range by table-look-up. First of all, we need to define a pixel-order for all pixels in the PRB searching range. Naturally, we use the order defined in the horizontally scanned string matching mode, that is, all pixels are ordered one LCU followed by another LCU in LCU coding order. Within an LCU, horizontal scanning is employed to order pixels. All pixels with the same hash-value are chained together, following the pixel-order. The hash-value of any current pixel being coded is calculated and the encoder only needs to search through the hash chain of the same hash-value, instead of all pixels of the entire PRB searching range, to find a potential matching reference pixel. This potential pixel is the first pixel of a potential reference string. In this way, the searching time can be significantly reduced.

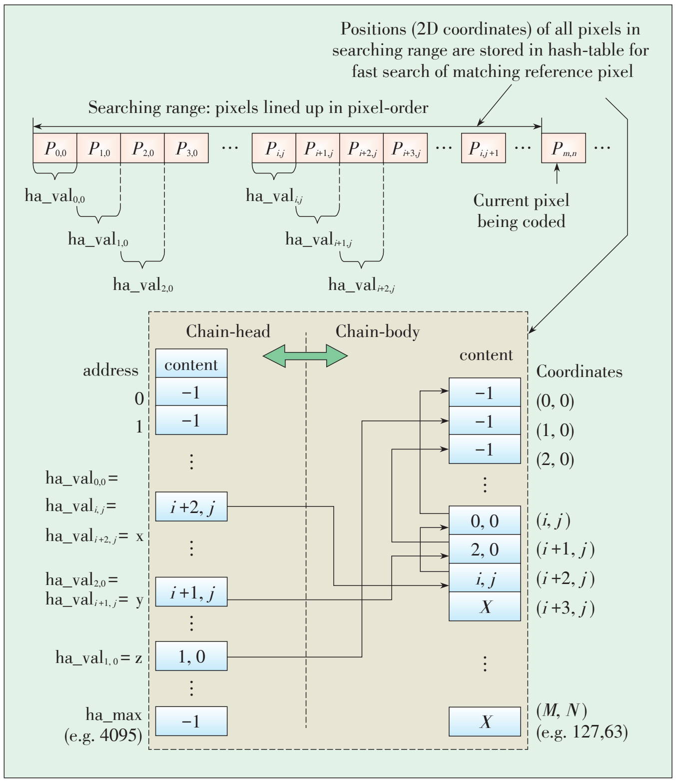
In **Fig. 4**, the pixels (in the searching range) lined up in pixel-order are $P_{0,0}$, $P_{1,0}$, $P_{2,0}$, $P_{3,0}$, ... , $P_{i,j}$, $P_{i+1,j}$, $P_{i+2,j}$, $P_{i+3,j}$, ... , $P_{i,j+1}$, ... , followed by the current pixel being coded, $P_{m,n}$. The hash values ha_val$_{0,0}$, ha_val$_{1,0}$, ha_val$_{2,0}$, ha_val$_{3,0}$, ... , ha_val$_{i,j}$, ha_val$_{i+1,j}$, ha_val$_{i+2,j}$, ha_val$_{i+3,j}$, ... , ha_val$_{i,j+1}$ are computed from the corresponding pixels. There are three hash chain examples in this figure.

1) Example 1

Three pixels $P_{0,0}$, $P_{i,j}$, and $P_{i+2,j}$ have the same hash value, i.e. ha_val$_{0,0}$ = ha_val$_{i,j}$ = ha_val$_{i+2,j}$ = x. Coordinates (0, 0), (i, j), and (i+2, j) of the three pixels are stored in the hash-table and form a hash chain of hash-value x. The hash-table actually has two parts: chain-head and chain-body. Among the three coordinates, the coordinates (i+2, j) having the largest pixel-order is stored in the chain-head slot of address x. The coordinates (i, j) having the second largest pixel-order is stored in chain-body location of coordinates (i + 2, j). The coordinates (0, 0) having the third largest pixel-order is stored in the chain-body location of coordinates (i, j). Finally, -1 is stored in the chain-body location of coordinates (0, 0) to indicate the end of the chain.

2) Example 2

Two pixels $P_{2,0}$ and $P_{i+1,j}$ have the same hash value, i.e. ha_val$_{2,0}$ = ha_val$_{i+1,j}$ = y. Coordinates (2, 0) and (i+1, j) form a hash chain of hash-value y in the hash-table. In the hash chain, the coordinates (i+1, j) having the largest pixel-order is stored in the chain-head slot of address y. The coordinates (2, 0) having the second largest pixel-order is stored in the chain-body location of coordinates (i + 1, j). Finally, -1 is



▲**Figure 4. Hash -table structure and contents.**

stored in the chain-body location of coordinates (2, 0) to indicate the end of the chain.

3) Example 3

The third hash chain has only one pixel $P_{1,0}$ whose hash value is ha_val$_{1,0}$ = z. Therefore, the coordinates (1, 0) is stored in the chain-head slot of address z and -1 is stored in the chain-body location of coordinates (1, 0) to indicate the end of the chain.

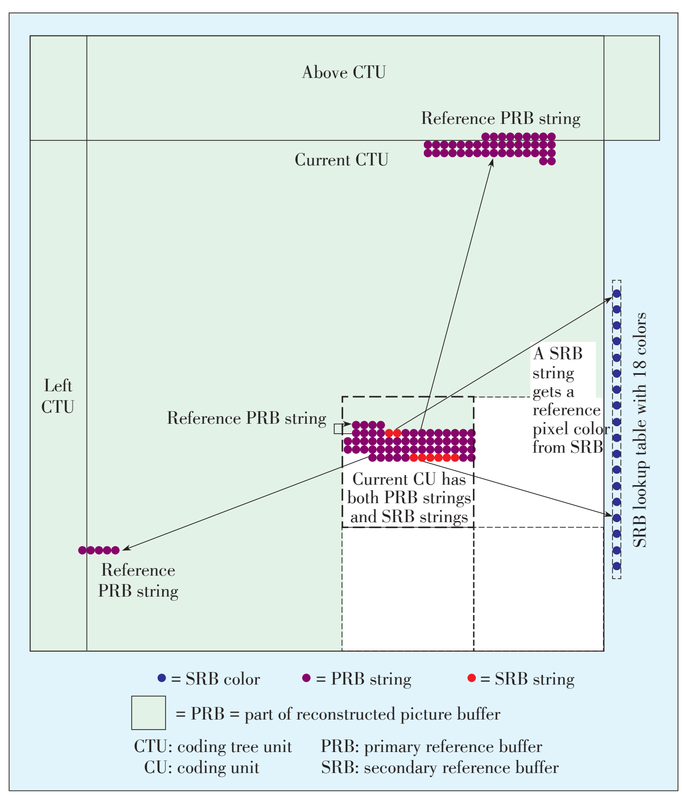
When the current pixel $P_{m,n}$ is encoded and the current string starts from $P_{m,n}$, the hash value ha_val$_{m,n}$ of $P_{m,n}$ is first computed and then used as the chain-head address to find the 1st coordinates in the hash chain of the same hash value. In the hash chain examples in 0, if ha_val$_{m,n}$ = x, the 1st coordinates is (i+2, j); If ha_val$_{m,n}$ = y, the 1st coordinates is (i+1, j); If ha_val$_{m,n}$ = z, the 1st coordinates is (1, 0). The content of the 1st coordinates in the chain-body is the 2nd coordinates in the hash chain of the same hash value, and the content of the 2nd coordinates in the chain-body is the 3rd coordinates in the hash chain of the same hash value, and so on. Therefore, the string-matching encoder can use the hash value of the current pixel being coded to find all pixels and their locations that have the same hash value in the searching range. Moreover, a hash chain starts from a hash-head slot, the 1st coordinates have the largest pixel-order, the 2nd coordinates have the second largest pixel-order, and so on.

For a pixel P = (Y, U, V) or (G, B, R), where Y, U, V (or G, B, R) are three 8-bit color components of P, a 12-bit hash value ha_val is computed. For lossy coding, ha_val is computed by concatenating 4-bit most significant bit (MSB) of Y, U, V (or G, B, R) to form a 12-bit value. For lossless coding, ha_val is computed by cyclic redundancy check (CRC)-12 algorithm [34] to get a 12-bit value. Obviously, all the pixels are always located in the same hash chain, no matter they have the identical 4-bit MSB component value in a lossy coding case or the identical 8-bit component value in a lossless coding case. Therefore, a hash chain with the same hash value provides a small and efficient candidate set of reference pixels for the starting pixel of a potential reference string.

4 PRB and SRB Based String-Matching

In the proposed string-matching technique, a current string being coded gets reference pixels from either PRB or SRB. When a current string gets the reference pixels from PRB, the reference pixels form a reference string that has exactly the same 2D shape and length (number of pixels) as the current string. The reference string can be in any valid location inside PRB. When a current string gets the reference pixels from SRB, it actually gets only one single reference pixel color from SRB LUT and all pixels of the entire string has the same color.

A CU coded by the string-matching technique can have both PRB and SRB strings (**Fig. 5**). In the figure, the 1st string in purple dots is a 4-pixel PRB string. Its reference string is just above it and has the same horizontal line shape and length of 4



▲**Figure 5. A CU coded by string -matching technique.**

pixels. The 2nd string in red dots is a 2-pixel SRB string. Its reference pixels are the 1st SRB LUT pixel color duplicated twice. The 3rd string in purple dots is a 43-pixel PRB string. Its reference string is located across the boundary between current coding tree unit (CTU) and above CTU. The PRB string and its reference string have exactly the same horizontal traverse-scan shape and length of 43 pixels. The 4th string in red dots is a 6-pixel SRB string. Its reference pixels are the 15th SRB LUT pixel color duplicated six times. The 5th string in purple dots is a 5-pixel PRB string. Its reference string is located across the boundary between current CTU and left CTU. The PRB string and its reference string have exactly the same horizontal line shape and length of 5 pixels.

In a string-matching encoder, the best matching reference string for a current starting pixel $P_{m,n}$ being coded is found via the following steps:

1) Searching the best reference PRB string for a current string cur$S_{m,n}$ via three sub-steps:

a) Calculating the hash value ha_val$_{m,n}$ of $P_{m,n}$.

b) For the pixel coordinates (i, j) obtained from the hash chain with the same hash value ha_val$_{m,n}$, determining the longest matching reference string ref$S_{i,j}$ = $\{S_{i,j}(k): 0 \leqslant k < L_{i,j}\}$, i.e. determining the largest $L_{i,j}$ that satisfies the constraint $|S_{i,j}(k)_Y - S_{m,n}(k)_Y| \leqslant E$, $|S_{i,j}(k)_U - S_{m,n}(k)_U| \leqslant E$, $|S_{i,j}(k)_V - S_{m,n}(k)_V| \leqslant E$ for all $k < L_{i,j}$, where E is a predetermined threshold. After going through all the pixel coordinates (i, j) on the hash chain of hash value ha_val$_{m,n}$, multiple matching reference strings are found as best reference PRB string candidates.

c) The best reference PRB string is selected from the best reference PRB string candidates, based on average RD cost evaluation. For a given current string cur$S_{m,n}$ and its reference string ref$S_{i,j}$ of length L, the average RD cost is calculated by (1)

$$\text{avgRDcost}(\text{cur}S_{m,n}, \text{ref}S_{i,j}) = [\lambda \text{bits}(\text{ref}S_{i,j}) + \text{distortion}(\text{cur}S_{m,n}, \text{ref}S_{i,j})]/L \quad (1)$$

where bits(ref$S_{i,j}$) is the number of bits for coding the reference string ref$S_{i,j}$, λ is a weighting factor, and distortion(cur$S_{m,n}$, ref$S_{i,j}$) is the distortion between the current string cur$S_{m,n}$ and the reference string ref$S_{i,j}$.

2) Searching the best reference SRB string, which is straightforward because a reference SRB string is a simple duplication of an SRB pixel color.

3) Either the best reference PRB string or the best reference SRB string is selected as the best matching reference string based on average RD cost evaluation.

If no PRB string or SRB string can be found for the current pixel $P_{m,n}$ being coded, the pixel itself is coded directly as an unmatched pixel.

5 Experimental Results

As an implementation example, the proposed string-matching technique is integrated into HM-16.4+SCM-4.0 reference software [31]. All experimental results were generated under the common test conditions and configurations defined in [30] for HEVC SCC project.

Thirteen test sequences (**Table 1**) are used in the experiment. The test sequences are classified into four categories: text and graphics with motion (TGM), mixed content (MC), camera captured (CC), and animation (ANI). Each test sequence has a RGB color format version and a YCbCr (YUV) color format version. Therefore, there are 26 sequences in total used in the experiment.

To evaluate the overall coding performance, the HEVC BD-rate metric [32], [33] is used for lossy coding and bit-rate saving metric is used for lossless coding. In lossy coding, an average BD-rate reduction is calculated by three color components G/Y, B/U, and R/V for each color format and category. In lossless coding, four bit-rate saving values (total, average, minimum, and maximum) are calculated for each color format and category. Both lossy coding and lossless coding experiments used three configurations: all intra (AI), random access (RA), and low delay B (LB). Encoding and decoding software runtime were also compared for evaluating the complexity of the encoder and decoder.

Two coding methods were compared:

1) HM-16.4+SCM-4.0 (SCM) with default setting. In particular, the IBC search range is full frame.

2) HM-16.4+SCM-4.0 integrated with string-matching technique (SCM-SM). The string-matching search range is the current LCU and left LCU.

Table 2 shows coding performance comparison (BD-rate reduction percentage in negative numbers) between SCM and SCM-SM in the lossy coding case. **Table 3** shows coding performance comparison (bit-rate saving percentage in negative numbers) between SCM and SCM-SM in the lossless coding case. Each row of data in the two tables corresponds to a com-

Table 1. Four-category test sequences

Resolution	Sequence name	Category	fps	Frames to be encoded
1920×1080	sc_flyingGraphics_1920x1080_60_8bit	TGM	60	0-299*
	sc_desktop_1920x1080_60_8bit	TGM	60	0-599
	sc_console_1920x1080_60_8bit	TGM	60	0-599
	MissionControlClip3_1920x1080_60p_8b444	MC	60	0-599
	EBURainFruits_1920x1080_50_10bit	CC	50	0-249**
	Kimono1_1920x1080_24_10bit	CC	24	0-119***
1280×720	sc_web_browsing_1280x720_30_8bit	TGM	30	0-299
	sc_map_1280x720_60_8bit	TGM	60	0-599
	sc_programming_1280x720_60_8bit	TGM	60	0-599
	sc_SlideShow_1280x720_20_8bit	TGM	20	0-499
	sc_robot_1280x720_30_8bit	ANI	30	0-299
2560×1440	Basketball_Screen_2560x1440_60p_8b444	MC	60	322-621
	MissionControlClip2_2560x1440_60p_8444	MC	60	120-419

*Only the first 300 frames of this sequence are used.
**Only the first 250 frames of this 10 -bit sequence are used, and InternalBitDepth is set to 8
***Only the first 120 frames of this 10 -bit sequence are used, and InternalBitDepth is set to 8
ANI: animation
CC: camera -captured content
MC: mixed content
TGM: Text and graphics with motion

▼Table 2. Performance comparison between SCM and SCM-SM in the lossy coding case

Anchor: SCM	All intra			Random access			Low delay B		
Tested: SCM-SM	G/Y	B/U	R/V	G/Y	B/U	R/V	G/Y	B/U	R/V
RGB, text & graphics with motion, 1080p & 720p	−3.77%	−4.23%	−4.19%	−2.48%	−2.65%	−2.68%	−1.79%	−1.92%	−1.95%
RGB, mixed content, 1440p & 1080p	−1.15%	−1.68%	−1.70%	−0.62%	−1.01%	−1.03%	−0.15%	−0.95%	−0.77%
RGB, Animation, 720p	0.02%	−0.03%	−0.02%	0.07%	0.02%	0.08%	0.00%	0.05%	0.10%
RGB, camera captured, 1080p	0.03%	0.02%	0.03%	0.11%	0.07%	0.14%	0.06%	0.01%	0.04%
YUV, text & graphics with motion, 1080p & 720p	−4.19%	−4.31%	−4.33%	−2.38%	−2.38%	−2.64%	−1.72%	−1.95%	−1.85%
YUV, mixed content, 1440p & 1080p	−1.51%	−2.52%	−2.79%	−0.85%	−1.95%	−2.36%	−0.31%	−1.74%	−2.04%
YUV, Animation, 720p	0.01%	−0.02%	−0.03%	0.02%	−0.18%	0.17%	0.20%	0.18%	0.31%
YUV, camera captured, 1080p	0.04%	0.03%	0.04%	0.10%	−0.07%	0.04%	0.02%	0.03%	0.09%
Encoding time (%)	121.69%			111.08%			106.73%		
Decoding time (%)	100.04%			106.50%			107.87%		

SCM: HM-16.4+SCM-4.0 SCM-SM: HM-16.4+SCM-4.0 integrated with string-matching technique

▼Table 3. Performance comparison between SCM and SCM-SM in lossless coding case

Anchor:SCM	All intra				Random access				Low delay B			
Tested: SCM-SM	Bit-rate change (Total)	Bit-rate change (Avg)	Bit-rate change (Max)	Bit-rate change (Min)	Bit-rate change (Total)	Bit-rate change (Avg)	Bit-rate change (Max)	Bit-rate change (Min)	Bit-rate change (Total)	Bit-rate change (Avg)	Bit-rate change (Max)	Bit-rate change (Min)
RGB, TGM	−3.87%	−4.64%	−14.0%	−0.47%	−2.19%	−2.99%	−9.56%	−0.26%	−2.06%	−2.28%	−6.12%	−0.20%
RGB, MC	−0.63%	−0.69%	−1.19%	−0.16%	−0.11%	−0.10%	−0.11%	−0.09%	−0.06%	−0.06%	−0.08%	−0.04%
RGB, ANI	0.00%	0.00%	0.00%	0.00%	0.01%	0.01%	0.01%	0.01%	0.00%	0.00%	0.00%	0.00%
RGB, CC	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
YUV, TGM	−4.39%	−5.17%	−14.5%	−0.31%	−2.39%	−3.47%	−9.90%	−0.16%	−2.23%	−2.76%	−6.46%	−0.13%
YUV, MC	−0.73%	−0.78%	−1.31%	−0.15%	−0.12%	−0.11%	−0.14%	−0.08%	−0.05%	−0.06%	−0.07%	−0.04%
YUV, ANI	0.00%	0.00%	0.00%	0.00%	0.01%	0.01%	0.01%	0.01%	0.01%	0.01%	0.01%	0.01%
YUV, CC	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.01%
Encoding time (%)	152.10%				111.39%				107.15%			
Decoding time (%)	100.67%				106.44%				105.89%			

SCM: HM -16.4+SCM -4.0 SCM -SM: HM -16.4+SCM -4.0 integrated with string -matching technique

bination of one color format and one category. There are totally eight combinations. Each combination contains 1—7 sequences. The encoding and decoding runtime ratios are also shown in the tables.

The experimental results include:

1) In the lossy coding case, SCM-SM can achieve up to 4.33% for AI, 2.68% for RA, 2.04% for LB average BD-rate reduction over SCM.
2) In the lossless coding case, SCM - SM can achieve up to 14.5% for AI, 9.9% for RA, 6.46% for LB maximum bit-rate saving and 5.17% for AI, 3.47% for RA, 2.76% for LB average bit-rate saving over SCM.
3) The improvement of SCM-SM over SCM depends on the contents. In YUV TGM case, the average bit - rate saving is 5.17% in lossless coding case and the average BD-rate reduction of components Y, U and V are 4.19%, 4.31% and 4.33%, respectively in lossy coding case for AI configuration. SCM-SM also has good BD-rate reduction over SCM for mixed content in all configurations, but no coding performance improvement for camera captured and animation contents.
4) In the lossy coding case, encoding runtime increase is about 22% for AI, 11% for RA, 7% for LB. In lossless coding case, encoding runtime increase is about 52% for AI, 11% for RA, 7% for LB

6 Conclusions

This paper proposes a string-matching coding technique for SCC. Both PRB and SRB are used for string-matching. The resulting bitstream is a combination of PRB string coding parameters, SRB string coding parameters, and unmatched pixel mixed on a string-by-string basis. The experiments using common test condition [30] show that the string-matching coding technique can achieve a lossy coding BD-rate reduction of up to 4.33%, 2.68%, 2.04% for AI, RA, and LB configurations respectively, and a lossless coding average bit-rate saving of up to 5.17%, 3.47%, 2.76% for AI, RA, and LB configurations respectively.

Our future work includes: 1) implementing rate-control in string-matching enhanced coding system; 2) optimizing string-matching coding to improve coding performance and reduce coding complexity further; 3) achieving idempotent (re-loss-free) [35] coding in repetitive (multi-generation) compression using the string-matching enhanced coding system.

References

[1] T. Lin, K. Zhou, and S. Wang. "Cloudlet-screen computing: a client-server architecture with top graphics performance," *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 13, no. 2, pp. 96–108, Jun. 2013. doi: 10.1504/IJAHUC.2013.054174.

[2] Y. Lu, S. Li, and H. Shen, "Virtualized screen: a third element for cloud_mobile convergence," *IEEE Multimedia*, vol. 18, no. 2, pp. 4–11, Apr. 2011. doi: 10.1109/MMUL.2011.33.

[3] T. Lin and S. Wang. "Cloudlet-screen computing: a multi-core-based, cloud-computing-oriented, traditional-computing-compatible parallel computing paradigm for the masses," in *IEEE International Conference on Multimedia and Expo*, New York, USA, Jul. 2009, pp. 1805–1808. doi: 10.1109/ICME.2009.5202873.

[4] T. Lin, K. Zhou, X. Chen, and S. Wang, "Arbitrary shape matching for screen content coding," *Picture Coding Symposium*, San Jose, USA, Dec. 2013, pp. 369–372. doi: 10.1109/PCS.2013.6737760.

[5] W. Zhu, J. Xu, W. Ding, *et al.*, "Adaptive LZMA-based coding for screen content," *Picture Coding Symposium*, San Jose, USA, Dec. 2013, pp. 373–376. doi: 10.1109/PCS.2013.6737761.

[6] T. Lin, X. Chen, and S. Wang, "Pseudo-2-D-matching based dual-coder architecture for screen contents coding," *IEEE International Conference on Multimedia and Expo Workshops*, San Jose, USA, Jul. 2013, pp. 1–4. doi: 10.1109/ICMEW.2013.6618315.

[7] S. Wang and T. Lin, "Compound image compression based on unified LZ and hybrid coding," *IET Image Processing*, vol. 7, no. 5, pp. 484–499, May 2013. doi: 10.1049/iet-ipr.2012.0439.

[8] T. Lin, P. Zhang, S. Wang, *et al.*, "Mixed chroma sampling-rate high efficiency video coding for full-chroma screen content," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 1, pp. 173–185, Jan. 2013. doi: 10.1109/TCSVT.2012.2223871.

[9] W. Zhu, W. Ding, R. Xiong, *et al.*, "Compound image compression by multi-stage prediction," *IEEE Visual Communications and Image Processing Conference*, San Diego, USA, Nov. 2012, pp. 1–6. doi: 10.1109/VCIP.2012.6410758.

[10] S. Wang and T. Lin, "United coding method for compound image compression," *Multimedia Tools and Applications*, vol. 71, no. 3, pp. 1263–1282, Aug. 2014. doi: 10.1007/s11042-012-1274-y.

[11] S. Wang and T. Lin, "United coding for compound image compression," *3rd International Congress on Image and Signal Processing*, Yantai, China, pp. 566–570, Oct. 2010. doi: 10.1109/CISP.2010.5647270.

[12] S. Wang and T. Lin, "A unified LZ and hybrid coding for compound image partial-lossless compression," *2nd International Conference on Image and Signal Processing*, Tianjin, China, Oct. 2009, pp. 1–5. doi: 10.1109/CISP.2009.5301019.

[13] W. Ding, Y. Lu, and F. Wu, "Enable efficient compound image compression in h.264/AVC intra coding," *IEEE International Conference on Image Processing*, San Antonio, USA, Oct. 2007, vol. 2, pp. 337–340. doi: 10.1109/ICIP.2007.4379161.

[14] ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/16, "Joint call for proposals for coding of screen content," ISO/IEC JTC 1/SC 29/WG 11 (MPEG), San Jose, USA, Doc. N14175, Jan. 2014.

[15] D.-K. Kwon and M. Budagavi, "Intra motion compensation with variable length intra MV coding," JCT-VC, Vienna, Austria, JCTVC-N0206, Jul. 2013.

[16] C. Pang, J. Sole, L. Guo, *et al.*, "Intra motion compensation with 2-D MVs," JCT-VC, Vienna, Austria, JCTVC-N0256, Jul. 2013.

[17] C. Pang, J. Sole, L. Guo, *et al.*, "Displacement vector signaling for intra block copying," JCT-VC, Geneva, Switzerland, JCTVC-O0154, Oct. 2013.

[18] T. Lin, S. Wang, P. Zhang, and K. Zhou, "AHG7: full-chroma (YUV444) dictionary+hybrid dual-coder extension of HEVC," JCT-VC, Shanghai, China, JCTVC-K0133, Oct. 2012.

[19] M. Budagavi, "Cross-check of JCTVC-K0133 (dictionary+hybrid dual-coder extension of HEVC)," JCT-VC, Shanghai, China, JCTVC-K0329, Oct. 2012.

[20] T. Lin, S. Wang, P. Zhang, and K. Zhou, "AHG8: P2M based dual-coder extension of HEVC," JCT-VC, Geneva, Switzerland, JCTVC-L0303, Jan. 2013.

[21] J. Ye, S. Liu, S. Lei, *et al.*, "Improvements on 1D dictionary coding," JCT-VC, Valencia, Spain, JCTVC-Q0124, 2014.

[22] R. Cohen, "Crosscheck for JCTVC-Q0124 improvements on 1D dictionary coding mode," JCT-VC, Valencia, Spain, JCTVC-Q0125, 2014.

[23] L. Zhao, X. Chen, T. Lin, *et al.*, "SCCE4: Results of subtest 3.3," JCT-VC, Sapporo, Japan, JCTVC-R0060, 2014.

[24] B. Li and J. Xu, "Cross-check of test 3.3 (JCTVC-R0060)," JCT-VC, Sapporo, Japan, JCTVC-R0109, 2014.

[25] K. Zhou, L. Zhao, X. Chen, and T. Lin, "Non-CE10: improvement on coding of ISC parameters and comparison to Palette coding," JCT-VC, Strasbourg, France, JCTVC-S0250, Oct. 2014.

[26] L. Zhao, K. Zhou, and T. Lin, "CE3: results of test B.4.2 (minimum string length of 20 pixels) on intra string copy," JCT-VC, Geneva, Switzerland, JCTVC-T0136, Feb. 2015.

[27] C.-H. Hung, Y.-J. Chang, J.-S. Tu, *et al.*, "CE3: crosscheck of CE3 test B.4.2 (JCTVC-T0136)," JCT-VC, Geneva, Switzerland, JCTVC-T0179, Feb. 2015.

[28] L. Zhao, K. Zhou, S. Wang, and T. Lin, "Non-CE3: improvement on intra string copy," JCT-VC, Geneva, Switzerland, JCTVC-T0139, Feb. 2015.

[29] R.-L. Liao, C.-C. Chen, W.-H. Peng, and H.-M. Hang, "Crosscheck of non-CE3: improvement on intra string copy (JCTVC-T0139)," JCT-VC, Geneva, Switzerland, JCTVC-T0200, Feb. 2015.

[30] H. Yu, R. Cohen, K. Rapaka, and J. Xu, "Common conditions for screen content coding tests," JCT-VC, Strasbourg, France, JCTVC-S1015, Oct. 2014.

[31] Heinrich Hertz Institute. (2015). *Rec. ITU-T H.265 | ISO/IEC 23008-2 High Efficiency Video Coding* [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.4+SCM-4.0

[32] G. Bjøntegaard, "Calculation of average PSNR differences between RD-Curves," ITU-T SG16 Q.6 Document, VCEG-M33, Austin, USA, Apr. 2001.

[33] G. Bjøntegaard, "Improvements of the BD-PSNR model," ITU-T SG16 Q.6 Document, VCEG-AI11, Berlin, Germany, Jul. 2008.

[34] T. V. Ramabadran and S. S. Gaitonde, "A tutorial on CRC computations," *IEEE Micro*, vol. 8, no. 4, pp. 62–75, Aug. 1988. doi: 10.1109/40.7773.

[35] T. Lin, "Achieving re-loss-free video coding," *IEEE Signal Processing Letters*, vol. 16, no. 4, pp. 323–326, Apr. 2009. doi: 10.1109/LSP.2009.2014285.

Manuscript received: 2015-05-13

Biographies

Tao Lin (lintao@tongji.edu.cn) received his MS and PhD degrees from Tohoku University, Japan, in 1985 and 1989. He has been with VLSI Lab, Tongji University, China since 2003. In 2005, he was awarded "Chang Jiang Scholars", the highest honor given by China Ministry of Education. From 1988 to 2002, he was a postdoctoral researcher with University of California, Berkeley, and developed multimedia ICs and products at several companies in Silicon Valley, including Integrated Device Technology, Inc., PMC-Sierra Inc., Cypress Semiconductor Corp., and NeoMagic Corp. He has been granted 24 US patents and 14 China patents. His current research interests include cloud-mobile computing, digital signal processing, audiovisual coding, and multimedia SoC design.

Kailun Zhou (vlsi@tongji.edu.cn) received his MS degree from Shanghai Jiaotong University, China in 2003. He is currently pursuing the PhD degree with Tongji University, China. His current research interests include embedded system design, video coding, and ASIC architecture, design and verification.

Liping Zhao (vlsi@tongji.edu.cn) received her MS degree in computer science and technology from Hunan University, China in 2009. She is currently a PhD candidate in control science and engineering at VLSI lab of Tongji University, China. Her current research interests include screen content coding and video coding.

DOI: 10.3969/j. issn. 1673-5188. 2015. 04. 009
http://www.cnki.net/kcms/detail/34.1294.TN.20151123.1459.002.html, published online November 23, 2015

Predicting LTE Throughput Using Traffic Time Series

Xin Dong[1], **Wentao Fan**[1], **and Jun Gu**[2]
(1. Beijing University of Posts and Telecommunications, Beijing, 100876, China;
2. ZTE Corporation, Shanghai 201203, China)

Abstract

Throughput prediction is essential for congestion control and LTE network management. In this paper, the autoregressive integrated moving average (ARIMA) model and exponential smoothing model are used to predict the throughput in a single cell and whole region in an LTE network. The experimental results show that these two models perform differently in both scenarios. The ARIMA model is better than the exponential smoothing model for predicting throughput on weekdays in a whole region. The exponential smoothing model is better than the ARIMA model for predicting throughput on weekends in a whole region. The exponential smoothing model is better than the ARIMA model for predicting throughput in a single cell. In these two LTE network scenarios, throughput prediction based on traffic time series leads to more efficient resource management and better QoS.

Keywords

ARIMA; exponential smoothing method; throughput prediction

1 Introduction

In recent years, there is a trend towards users accessing the Internet from a variety of applications and without restriction in terms of geographic location. This has resulted in an exponential increase of wireless traffic. In 2012, global wireless data traffic grew 70 percent year on year [1]. Thus, mobile network operators have to make a use of limited resources to meet ever-increasing traffic demands. To plan and run networks efficiently, it is important to understand the statistical characteristics of data traffic by analyzing the real traffic.

In [2], the authors use the throughput measured from a real-work cellular network to statistically model time-varying throughput per cell and the distribution of instantaneous throughput per cell over different cells. The proposed statistical models can be used to simulate the time-varying and location-varying throughput of cells. In [3], the authors analyze several widely accepted throughput network-performance indicators in LTE. Their analysis is based on counters and call traces of a live network. However, neither [1] nor [2] describe a scenario where throughput in a whole region changes over time. In [4], the authors estimate this throughput using a formula that expresses the behavior of TCP throughput. We consider throughput data as a time series that can be predicted using data measured in the past.

In this paper, we consider two practical scenarios: whole region and single cell. In the first scenario, we constructed a better model than both the individual ARIMA model and exponential smoothing model for predicting downlink throughput on weekdays and weekends in a whole region. In the second scenario, the traffic load in a single cell is uncertain and varying over time. We construct a model for predicting the instantaneous downlink throughput in a single cell of a large urban cellular network.

2 Data Set and Modeling Methodology

2.1 Data Description

Our data set includes records of Internet downloads and uploads in Hong Kong. The data was collected from 1352 cell sites across the city over 21 days between February and March 2014. Each data session includes the throughput of the downlink and uplink, timestamp, and cell ID. Each cell ID is also associated with the GPS coordinates of the corresponding cell. In this paper, LTE throughput is modeled as a time series and then predicted using an ARIMA model and exponential smoothing method.

2.2 Time Series Analysis

Time series data is an important class of data. Any change of an attribute value as a function of time can be considered time series data. Such data may derive from the atmosphere, commodity production, geography, sensors, the stock market, or inventory control. The throughput data in an LTE network can also be viewed as a time series. Prediction of time series is based on the idea that historical data related to past behavior can be used to predict the future behavior.

2.2.1 ARIMA Model

The autoregressive integrated moving average (ARIMA) model was introduced by Box-Jenkins [5]. ARIMA (p, d, q) is an autoregressive moving average (ARMA) model based on differenced time series data. The original time series data is differenced on the order d to make the data stationary. A stationary time series can be modeled as an ARMA model of order (p, q), where p is the order of the AR process and q is the order of the MA process. ARMA-modeled current time series data is given by:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + b_1 e_{t-1} + b_2 e_{t-2} + \dots + b_q e_{t-q} + e_t \quad (1)$$

where $y_{t-1}, y_{t-2} \cdots y_{t-p}$ are the data at past time points, $e_{t-1}, e_{t-2} \cdots e_{t-2}$ are the errors at past time points, e_t is a present error (ARMA assumes this error is Gaussian-distributed), $a_1, a_2, \dots, a_p$ are the AR coefficients, and $b_1, b_2, \dots, b_q$ are the MA model coefficients [6].

ARIMA (*p*, *d*, *q*) modeling involves making the data stationary, then identifying suitable values for the model order, then predicting the time series data from the model.

2.2.2 Exponential Smoothing Model

Exponential smoothing is a trend-analysis and prediction method based on the moving average method. Exponential smoothing method has three main submethods—linear exponential smoothing, secondary exponential smoothing and cubic exponential smoothing—that differ in terms of smoothing times [7]–[8]. The most common of these methods is secondary exponential smoothing, given by:

$$F_t^{(1)} = \alpha Y_t^{(1)} + (1-\alpha) F_{t-1}^{(1)}, F_t^{(2)} = \alpha Y_t^{(1)} + (1-\alpha) F_{t-1}^{(2)} \quad (2)$$

$$Y_{t+m} = a_t + b_t m \quad (3)$$

$$a_t = 2F_t^{(1)} - F_t^{(2)} \quad (4)$$

$$b_t = \frac{\alpha}{1-\alpha}\left(F_t^{(1)} - F_t^{(2)}\right) \quad (5)$$

where $F_t^{(1)}$ is the smoothed value of period *t*, $F_t^{(2)}$ is the second smoothed value of *t*, $F_{t-1}^{(2)}$ is the second exponential smoothed value of $t-1$, and α is the smoothing factor [9].

2.3 Metrics

Root-mean-square error (RMSE) and R-squared are used to determine how well the model fits. RMSE represents the mean-squared error statistics of the output model. These statistics show the difference between the model's predictions and real values, i.e., the standard deviation of the residuals. The unit of measure is consistent with the original data. The RMSE is given by [10]:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2} \quad (6)$$

where $\hat{y}_i$ is the real value, and y_i is the predicted value.

R-squared [11] is the square of the correlation between the measured (empirical) value and the predicted value. A higher R-squared means a better-fitting model. The maximum R-squared value is 1. When the time series contains seasonal trends, a stationary R-squared statistic is better than a normal R-squared statistic.

In this paper, we use stationary R-squared as the evaluation index for data with obvious seasonal trends. We use RMSE as the evaluation index for data with no obvious seasonal trend, such as throughput data from a single cell.

3 Modeling and Results

Here, we analyze two practical scenarios. In the first scenario, each cell is divided into regions, and the throughput of an entire region is predicted. In the second scenario, the throughput of a single cell is predicted according to historical data.

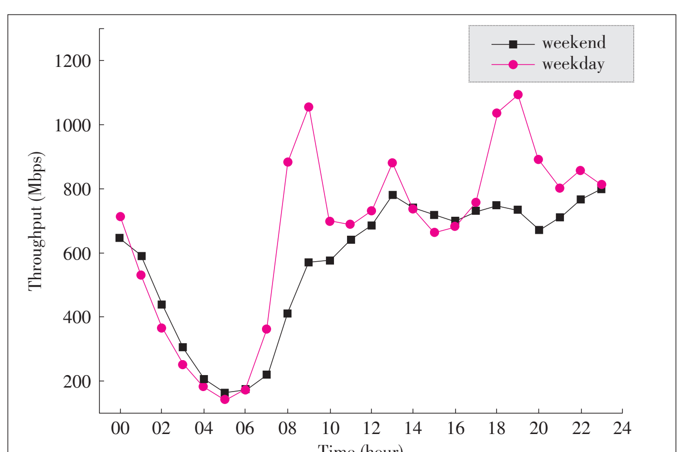
The reason for creating these two scenarios is that network operators are constantly constructing, adjusting, and optimizing their network, and single cell throughput prediction alone is not enough. If a new cell is built next to cell A, then the throughput of cell A is bound change, and the former data is discarded. Therefore, the first scenario is proposed. QoS can be improved by knowing the network throughput in advance.

3.1 Throughput Prediction for a Whole Region

We first investigate how downlink throughput in a whole region changes over time. **Fig. 1** shows the mean throughput in a region on weekdays and weekends. The weekday mean throughput was obtained by averaging the throughput in the whole region over 10 consecutive weekdays, and the weekend mean throughput was obtained by averaging the throughput over two consecutive weekends (four days). For both weekdays and weekends, the mean throughput in the whole region was at its lowest at 05:00. On a weekday, the mean throughput peaked at 09:00 and 19:00. On the weekend, throughput peaked at 13:00. We divided the throughput in the whole region in weekdays and weekends for further statistical analysis

To analyze the throughput on weekdays, we used the hourly data of ten consecutive weekdays. Five days of this data was used for modeling, and the other five days was used to determine the accuracy of the prediction.

In **Fig. 2**, the real throughput on weekdays in the whole re-



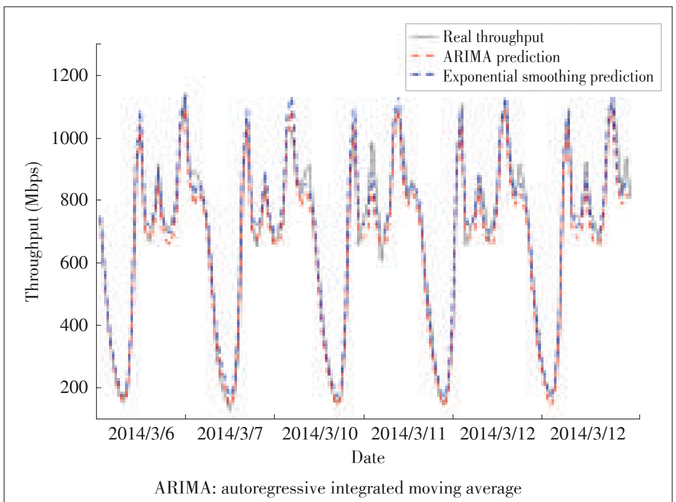
▲**Figure 1. Weekday and weekend mean throughput in a whole region over 24 hours.**

gion is seasonal. Therefore, we use the ARIMA (2, 0, 1) model and exponential smoothing with $\alpha = 0.600$ to predict throughput on weekdays in the whole region. Although there are gaps between the measured and predicted throughput in the whole region, the predictions by both models are highly accurate. The ARIMA model is more accurate in the valleys of the real throughput curve, which occur at around 05:00, 11:00 and 15:00 every weekday.

Table 1 shows the degree of fit statistics for the prediction models. Both the fit of the curve and the stationary R-squared statistic indicate that the ARIMA model is better than the exponential smoothing model for predicting throughput on weekdays in a whole region.

To study the throughput on weekends, we used hourly throughput data from two consecutive weekends. Two days of this data was used for modeling, and the other two days of data was used to determine the accuracy of the prediction.

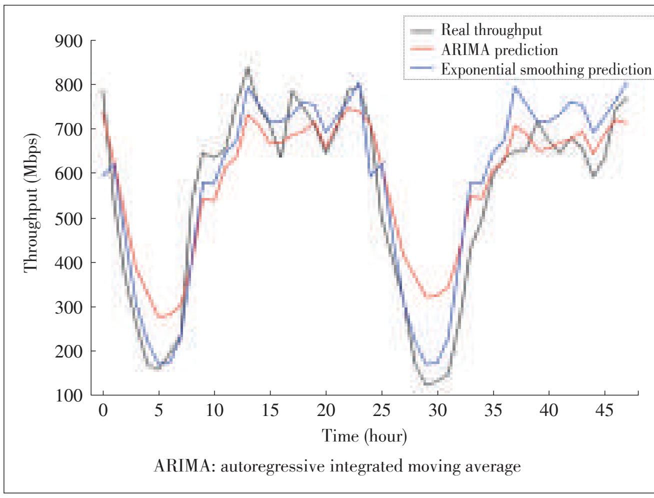
The prediction models for throughput of weekends in a whole region is ARIMA (1, 0, 2) and exponential smoothing method with $\alpha = 0.500$. **Fig. 3** shows predicted weekend throughput in a whole region using the ARIMA model and exponential smoothing model separately. The throughput predicted using the exponential smoothing model is closer to actual throughput that that predicted using the ARIMA model on a weekend in a whole region (**Table 2**). The degree of fit statistics supports this. Hence, we obtain the result, that exponential smoothing method is better to predict the weekends' throughput in a whole region.



▲Figure 2. ARIMA model and exponential smoothing model are used to predict the throughput on weekdays in a whole region.

▼Table 1. Degree of fit statistics for models used to predict throughput on weekdays in a whole region

	Stationary R-squared
ARIMA model	0.880
Exponential smoothing model	0.664
ARIMA: autoregressive integrated moving average	



▲Figure 3. ARIMA model and exponential smoothing model for predicting the throughput on weekends in a whole region.

▼Table 2. Degree of fit statistics for models used to predict throughput on weekends in a whole region

	Stationary R-squared
ARIMA model	0.460
Exponential smoothing model	0.723
ARIMA: autoregressive integrated moving average	

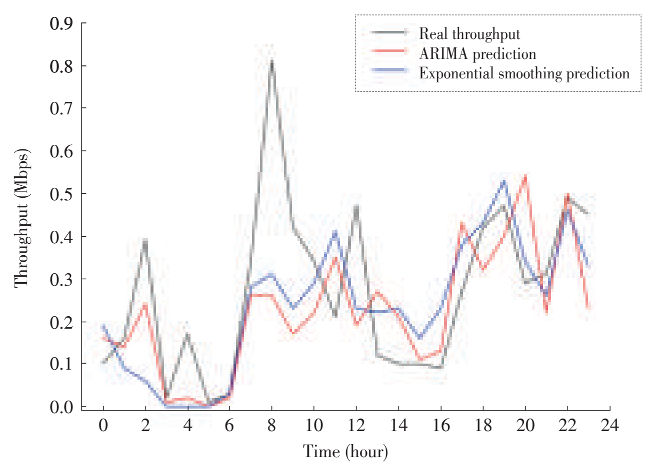
3.2 Throughput Prediction for a Single Cell

A single-cell traffic time series is highly unpredictable and has no obvious seasonal trend. Even within the same cell, throughput changes greatly on different days. Although there are gaps between the real and predicted throughput curves, a time series model for a single cell still has some use in network optimization. Here, we use the throughput data of an LTE network over eight consecutive days. Seven days of this data is used for modeling, and the other day of data is used to determine how well the model fits.

The stationary R-squared statistic is usually used as an evaluation index when the time series contains seasonal trends. Because there is no significant seasonal trend in the throughput of a single cell, we use RMSE as an evaluation index.

Fig. 4 shows the throughput prediction for single cell. The prediction models are ARIMA (1, 1, 1), and exponential smoothing with $\alpha = 0.100$. Fig. 4 shows that these two models do not accurately predict abrupt changes of throughput in the single cell. The exponential smoothing model is a little more accurate between 17:00 and 23:00. **Table 3** shows the accuracy statistics of the two models.

We chose 100 cells randomly and modeled them. Then we obtained the RMSE statistics for these cells. **Fig. 5** shows the distribution of RMSE for prediction using the ARIMA model and exponential smoothing model in 100 cells. The RMSE of



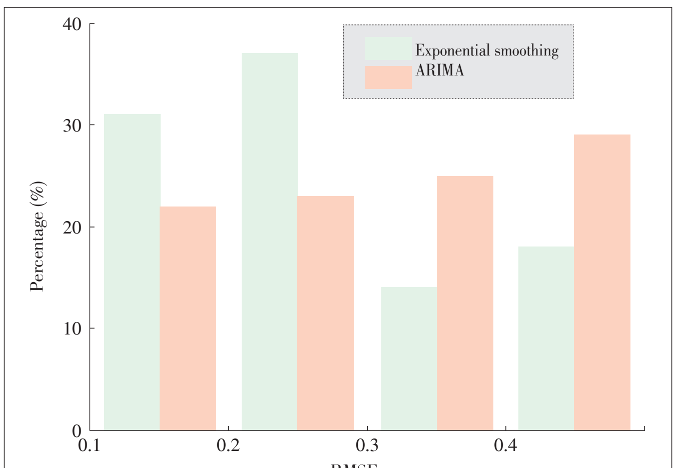
ARIMA: autoregressive integrated moving average

▲**Figure 4. ARIMA model and exponential smoothing method for predicting the throughput of a single cell in an LTE network.**

▼**Table 3. RMSE of the prediction model (single-cell throughput)**

	RMSE
ARIMA model	0.193
Exponential smoothing model	0.153

ARIMA: autoregressive integrated moving average
RMSE: root-mean-square error



ARIMA: autoregressive integrated moving average
RMSE: root-mean-square error

▲**Figure 5. RMSE statistics for throughput prediction in 100 cells.**

the exponential smoothing method is mainly distributed between 0 and 0.3, and that for the ARIMA model is mainly distributed above 0.3. In general, the exponential smoothing model is better for predicting throughput in a single cell.

4 Conclusion

In this paper, LTE throughput is modeled as a time series, and future values of the traffic time series are predicted using the ARIMA model and exponential smoothing model. Using different time series models, we studied throughput in both a single cell and a whole region within an LTE network. When studying throughput in a whole region, we considered weekday and weekend separately because their throughput patterns were different. The ARIMA model is better than exponential smoothing for predicting throughput on weekday in a whole region, and exponential smoothing model is much better than the ARIMA model for predicting throughput on weekends in a whole region. Exponential smoothing is more accurate than the ARIMA model for predicting throughput in a single cell. Throughput prediction based on time series models can be used in the design, management, planning, and optimization of networks.

References

[1] C. V. N. Index, "Global mobile data traffic forecast update, 2012-2017," Cisco White Paper, 2013.

[2] E. Nan, X. Chu, W. Guo, and J. Zhang, "User data traffic analysis for 3G cellular networks," in *Proc. CHINACOM*, Guilin ,China, Aug. 2013, pp. 469–472. doi: 10.1109/ChinaCom.2013.6694641.

[3] V. Buenestado, J. Ruiz-Aviles, M. Toril, *et al.*, "Analysis of Throughput Performance Statistics for Benchmarking LTE Networks," *IEEE Communications Letters*, vol. 18, no. 9, pp. 1607–1610, Sept. 2014.

[4] M. Mirza, J. Sommers, P. Barford, and X. Zhu, "A machine learning approach to TCP throughput prediction," in *Proc. SIGMETRICS*, New York, USA, 2007, pp. 97–108.

[5] G. E. P. Box and G. M. Jenkins, *Time Series Analysis Forecasting and Control*, 2nd ed. San Francisco, CA: Holden-Day, 1976.

[6] C. Babu and B. Reddy, "Predictive data mining on average global temperature using variants of ARIMA models", in *Proc. ICAESM*, Tamil Nadu, India, 2012, pp. 256–260.

[7] Q. Chen and X. Li, *System Engineering-Theory and Practice*. Beijing, China: National Defense Industry Press, 2009.

[8] X. Shang, W. Lin, and Y. Tang. "Development and application of a combined water quality prediction model based on exponential smoothing and GM(1,1)," *Environmental Science&Technology*, vol. 34, no. 1, pp. 191–195, May 2011, doi: 10.3969/j.issn.1003-6504.2011.01.046.

[9] W. Sun and R. Yang, *Economic Forecast*. Beijing, China: Agricultural University Press, 2005.

[10] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, pp. 679–688, 2006.

[11] J. R. Taylor, *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. Mill Valley, USA: Univ. Science Books, 1996.

Manuscript received: 2015-07-31

Biographies

Xin Dong (dongxin2014@gmail.com) is pursuing her master's degree in telecommunications at Beijing University of Post and Telecommunications (BUPT). Her research interests include data mining and time series analysis. She has previously researched the prediction of time serials of traffic flow.

Wentao Fan (ffantastic@126.com) is pursuing his master's degree in telecommunications at BUPT. His research interests include data mining, and network analysis and optimization based on mobile devices. He has researched the prediction of time serials of traffic flow using the SVR method.

Jun Gu (gu.jun@zte.com.cn) is a chief engineer of 4G radio network planning at ZTE Corporation. He has 10 years' research and field experience in network principles, standardization, simulation, algorithm design, and planning and optimization.

ZTE Communications

Table of Contents, Volume 13, Numbers 1–4, 2015

Volume–Number–Page

SPECIAL TOPICS

5G Wireless: Technology, Standard and Practice

Using Artificial Intelligence in the Internet of Things

Recent Advances in Smart Grid

Smart City: Key Technologies and Practices

ZTE Communications

Table of Contents, Volume 13, Numbers 1–4, 2015

REVIEW

RESEARCH PAPERS

ZTE Communications Guidelines for Authors

- **Remit of Journal**

ZTE Communications publishes original theoretical papers, research findings, and surveys on a broad range of communications topics, including communications and information system design, optical fiber and electro-optical engineering, microwave technology, radio wave propagation, antenna engineering, electromagnetics, signal and image processing, and power engineering. The journal is designed to be an integrated forum for university academics and industry researchers from around the world.

- **Manuscript Preparation**

Manuscripts must be typed in English and submitted electronically in MS Word (or compatible) format. The word length is approximately 4000 to 7000, and no more than 6 figures or tables should be included. Authors are requested to submit mathematical material and graphics in an editable format.

- **Abstract and Keywords**

Each manuscript must include an abstract of approximately 150 words written as a single paragraph. The abstract should not include mathematics or references and should not be repeated verbatim in the introduction. The abstract should be a self-contained overview of the aims, methods, experimental results, and significance of research outlined in the paper. Five carefully chosen keywords must be provided with the abstract.

- **References**

Manuscripts must be referenced at a level that conforms to international academic standards. All references must be numbered sequentially in-text and listed in corresponding order at the end of the paper. References that are not cited in-text should not be included in the reference list. References must be complete and formatted according to ZTE Communications Editorial Style. A minimum of 10 references should be provided. Footnotes should be avoided or kept to a minimum.

- **Copyright and Declaration**

Authors are responsible for obtaining permission to reproduce any material for which they do not hold copyright. Permission to reproduce any part of this publication for commercial use must be obtained in advance from the editorial office of *ZTE Communications*. Authors agree that a) the manuscript is a product of research conducted by themselves and the stated co-authors, b) the manuscript has not been published elsewhere in its submitted form, c) the manuscript is not currently being considered for publication elsewhere. If the paper is an adaptation of a speech or presentation, acknowledgement of this is required within the paper. The number of co-authors should not exceed five.

- **Content and Structure**

ZTE Communications seeks to publish original content that may build on existing literature in any field of communications. Authors should not dedicate a disproportionate amount of a paper to fundamental background, historical overviews, or chronologies that may be sufficiently dealt with by references. Authors are also requested to avoid the overuse of bullet points when structuring papers. The conclusion should include a commentary on the significance/future implications of the research as well as an overview of the material presented.

- **Peer Review and Editing**

All manuscripts will be subject to a two-stage anonymous peer review as well as copyediting, and formatting. Authors may be asked to revise parts of a manuscript prior to publication.

- **Biographical Information**

All authors are requested to provide a brief biography (approx. 150 words) that includes email address, educational background, career experience, research interests, awards, and publications.

- **Acknowledgements and Funding**

A manuscript based on funded research must clearly state the program name, funding body, and grant number. Individuals who contributed to the manuscript should be acknowledged in a brief statement.

- **Address for Submission**

magazine@zte.com.cn
12F Kaixuan Building, 329 Jinzhai Rd, Hefei 230061, P. R. China

ZTE COMMUNICATIONS

ZTE Communications has been indexed in the following databases:

- Cambridge Scientific Abstracts (CSA)
- China Science and Technology Journal Database
- Chinese Journal Fulltext Databases
- Inspec
- Norwegian Social Science Data Services (NSD)
- Ulrich's Periodicals Directory
- Wanfang Data—Digital Periodicals