http://www.icact.org/journal.asp

ICACT-TACT JOURNAL

Transactions on Advanced Communications Technology

Volume 3 Issue 3, Jan 2014, ISSN: 2288-0003

Editor-in-Chief
Prof. Thomas Byeongnam YOON, PhD.

Global IT Research Institute

Volume. 3 Issue. 1

Ecosystem-Driven Design of In-Home Terminals Based on Open Platform for the Internet-of-Things

Zhibo Pang*, Junzhe Tian**

* *ABB AB, Corporate Research, Forskargränd 7, 72178, Västerås, Västmanland, Sweden*

** *Royal Institute of Technology, Department of Electronic Systems, Isafjordsgatan 39, 16440, Kista, Sweden*

pang.zhibo@se.abb.com, junzhe@kth.se

***Abstract*—In-home healthcare services based on the Internet-of-Things (IoT) have great business potentials. To turn it into reality, a business ecosystem should be established first. Technical solutions should therefore aim for a cooperative ecosystem by meeting the interoperability, security, and system integration requirements. In this paper, we propose an ecosystem-driven design strategy and apply it in the design of an open-platform-based in-home healthcare terminal. A cooperative business ecosystem is formulated by merging the traditional healthcare and mobile internet ecosystems. To support the ecosystem in practical technology and business development, ecosystem-driven standardization efforts, security mechanisms, terminal design principles, and data handling schemes are analyzed and corresponding solutions or guidelines are presented. Thirdly, to verify the proposed design strategy and guidelines, an open-platform-based terminal is implemented and demonstrated by a prototyping system.**

***Index Terms*—Ecosystem-Driven Design; Internet-of-Things (IoT); In-Home Healthcare Station (IHHS); Open Platform;**

I. Introduction

THE revolution of Internet-of-Things (IoT) is reshaping the modern healthcare with promising economic and social prospects [1-3]. Powered by its ubiquitous identification, sensing, and communication capacities, all objects in the healthcare systems (people, equipment, medicine, etc.) can be tracked and/or monitored on a 24/7 basis [4]. Enabled by its global connectivity, all the healthcare related information (logistics, diagnosis, therapy, recovery, medication, management, finance, and even daily activity) can be collected, managed, and shared efficiently. By using the personal computing devices (laptop, mobile phone, tablet, etc.) and mobile internet access (WiFi, 3G, LTE, etc.), the IoT-based healthcare services can be mobile and personalized [5-7]. Large user base and maturated ecosystem of traditional mobile internet service have significantly sped up the development of the IoT-powered in-home healthcare (IHH) services, so-called Health-IoT. At the same time, the Health-IoT extends the traditional mobile internet services to a new application area. Especially after the open-source operation systems, such as Android [8], were introduced and broadly applied, the Health-IoT has been expected to be one of the "killer" applications of the IoT. Therefore the development of Health-IoT solution based on open platform has become a hot topic.

In recent years, a number of single point devices and applications have been developed. But as required by the economy of scale, a generic architecture is needed to support various applications by a common IoT platform. So, more comprehensive study is needed. Moreover, this general architecture should be feasible not only from technical point-of-view but also from business point-of-view. Comparing to the traditional mobile internet ecosystem, the Health-IoT ecosystem is much more complicated as more stakeholders are involved. To create sustainable Health-IoT services, the establishment of a cooperative ecosystem is primarily important to the whole industry. Such ecosystem should deliver enough added values to all stakeholders instead of a part. High level architectures of all technical aspects such as security, interoperability, and enterprise information system (EIS) integration, should serve for this goal. Therefore, ecosystem-driven design strategy is necessary in the early stage of technical development. The exiting research on this topic is very rare.

In this paper, extending our previous works in [14-18], an ecosystem-driven design strategy for the Health-IoT applications is presented and demonstrated.

First, a cooperative ecosystem of Health-IoT is formulated by Value Chain Analysis of the traditional healthcare and mobile internet ecosystems. The two traditional business ecosystems are destructed and merged into one new ecosystem.

Second, as the ecosystem of Health-IoT is established upon shared infrastructures, the interoperability of devices from

Manuscript received July 29, 2013. This work was supported in part by the VINN Excellence Center of iPack at the Royal Institute of Technology.

Zhibo Pang is with ABB AB, Corporate Research, Forskargränd 7, 72178, Västerås, Västmanland, Sweden (phone: +46-21 325104; fax: +46-21 323212; e-mail: pang.zhibo@se.abb.com).

Junzhe Tian is with Royal Institute of Technology, Department of Electronic Systems, Isafjordsgatan 39, 16440, Kista, Sweden. (e-mail: junzhe@kth.se).

different suppliers is important. By reviewing existing standardization efforts on device interoperability, we propose a set of simplified interfaces among different actors within the ecosystem.

Third, to support fair distribution of benefits among all stakeholders, value-centric security schemes are proposed, including the public authority-based authentication, the secure element (SE) based cryptography, and the non-invasive message handover.

Forth, in order to achieve the economy of scale, an IHH Station (IHHS) is proposed as a universal platform for device and service integration and convergence. Design principles of the open-platform-based terminals are detailed.

Fifth, ecosystem-driven data handling schemes are introduced including layered data compression and self-contained data formatting.

Finally, to verify the concepts and technical feasibilities, we have developed a prototype system called iMedBox. It is a specific case for medication management and in-home monitoring applications. The iMedBox hardware, software and backbone system are implemented and evaluated by field demonstrations. The positive feedbacks have proven the feasibility of proposed design methods, proposed architectures and solutions. Based on the results of this paper, economically feasible services are closer to reality.

The rest of this paper is organized as follows. The ecosystem analysis is presented in section II. The standardization efforts are reviewed in section III. The security schemes are presented in section IV. The design principles of open-platform-based terminals are given in section V. The data handling mechanisms are introduced in section VI. The implementation of the prototyping system and experimental results are discussed in section VII, and concluded in section VIII.

II. Ecosystem Reconstruction

A. *Lessons from the failure of Google Health*

Since Jan 1, 2012, as one of the most famous Health-IoT business efforts, the Google Health service has been discontinued [9]. This has been looked as a big setback. It is difficult to assert the exact reason but we can learn some lessons by analyzing the possible reasons. According to the summary of possible reasons listed by Brian Dolan [10], seven of the ten reasons are related to the establishment of ecosystem: *the Google Health was not trustworthy (lack of public authority), not fun or social, not involving doctors, not partnering with insurance companies, hard to overcome the current reimbursement barriers, lack of advertising opportunity, and not useful to consumers.*

This finding is consistent with the prediction of ITU when the vision of IoT was introduced: "*the Internet of Things will occur within a new ecosystem that will be driven by a number of key players*" [11]. Before developing the technical solutions, it is more important clearly answer "*how to establish a new cooperative ecosystem, and how to deliver enough added values to all of stakeholders in that ecosystem?*" Hence, the ecosystem analysis is the first step of our work.

B. *Ecosystems of traditional healthcare and mobile internet*

As shown in Fig. 1 (a) and (b), the ecosystems of traditional healthcare service and traditional mobile internet service are formularized and compared. The main stakeholders involved in both of them can be classified into four roles: financial sources, means suppliers, service providers, and end users. The service providers are the actor of service execution and delivery. Means providers provide necessary materials, tools, supplies, etc. to the service providers but seldom face the end users directly. Products and services mainly flow from means providers, through service providers, to end users. Payments (obligatory or optional, depending on different cases) flow back from end users, through financial sources, to the means providers and service providers. Thus a close-loop value chain is established. It is exactly the "close-loop" feature that makes the ecosystem economically sustainable. Win-win cooperation is enabled only if every stakeholder's benefit is guaranteed.

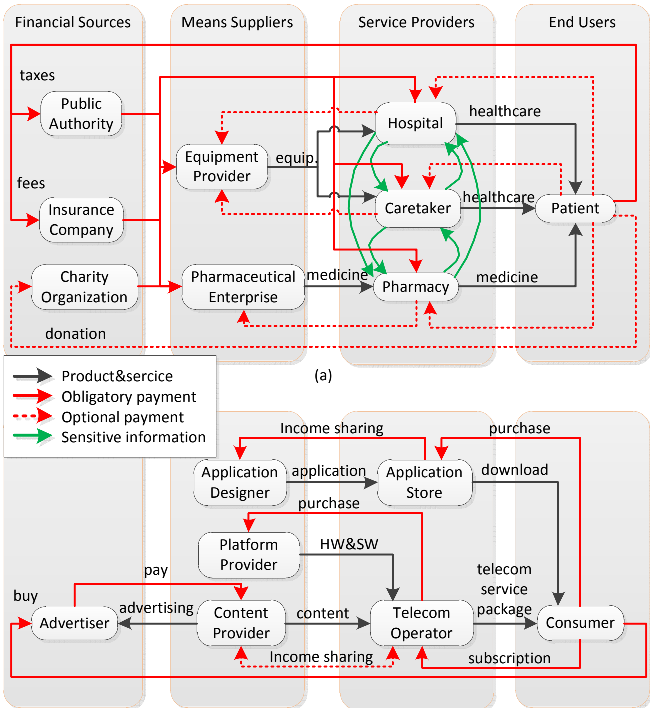


Fig. 1. Business ecosystems of (a) traditional healthcare service, (b) traditional mobile internet service

In spite of the above mentioned similarity, we can see significant differences between the two ecosystems. Firstly, the healthcare service ecosystem has more complicated financial sources. Despite the diverse policies in different countries, the public authority and insurance company are the most important financial sources, and thus have the highest influence on the rules of healthcare services. Another important difference is related to privacy and security. The healthcare services deal with privacies of end users which are much more sensitive than that in the mobile internet services. As a result, in the traditional healthcare ecosystem, the privacy information flows within the service providers strictly limited by regulations that have been well established and accepted. These two major differences are the main concerns and drivers when we formulate the new Health-IoT ecosystem.

C. *The proposed Health-IoT ecosystem*

As shown in Fig. 2 the new Health-IoT ecosystem is proposed and formulated by merging the two traditional ecosystems. Obviously the Health-IoT service is a business established on shared infrastructures including the internet backend facilities, core networks, access networks, and mobile terminals.

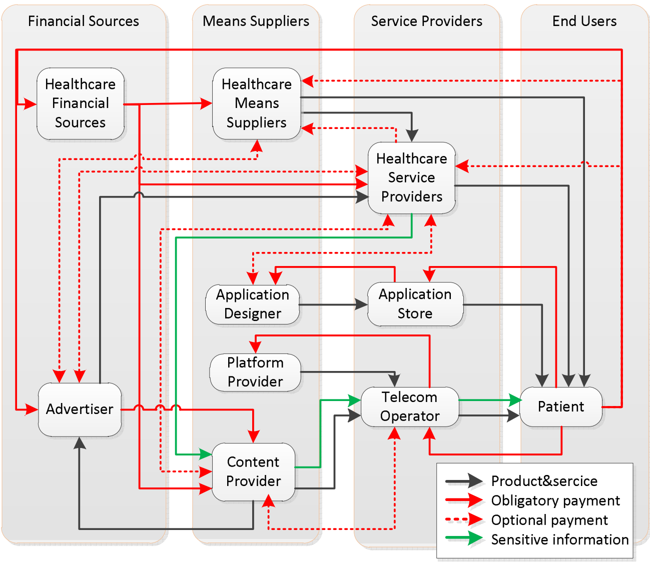


Fig. 2. The proposed cooperative ecosystem of the Health-IoT service

In this ecosystem, the healthcare service providers (like hospitals, elderly houses, pharmaceutical enterprises etc.) and healthcare financial sources (like public authorities, insurance companies, etc.) have larger influence than other stakeholders. The content providers (like Google, Amazon, Facebook, etc.) and telecom operators (like China Mobile, Vodafone, Verizon, etc.) cannot rule the ecosystem anymore. Large mobile device providers (like Apple, Nokia, Samsung, etc.) and medical device providers (like Roche, Omron, Philips, Johnson, etc.) should cooperate more than before to ensure the interoperability of their products. Due to the application-store-based software distribution model, consumers and application developers get more fairness in the ecosystem.

The cooperation between traditional healthcare service providers and internet content providers is the key to bring the ecosystem into reality. On one hand, the healthcare service providers don't need to establish new extra infrastructures (like data centers, servers, software and other backend systems) by their own. Instead, they should make use of the existing infrastructures owned by the internet content providers. In this case, the contents of healthcare services are delivered to the end users through the channels of telecom operators. On the other hand, the internet content providers, as well as telecom operators, should get the healthcare contents from healthcare service providers rather than "create" such contents by themselves. The healthcare financial sources should encourage and protect such cooperation by paying to the content providers directly or through healthcare service providers.

Furthermore, the privacy regulations and public authentications should be applied to the content providers and telecom operators, as strictly as they are applied to the healthcare service providers. This is the primary precondition for the end users to agree on uploading and managing their privacy through these channels. Besides the legislative approaches, technical approaches should also be in place to make sure only the owner and specially authorized individuals can access the private information. These principles are the foundation of the proposed security schemes.

Additionally, the advertisers should be authorized to provide specific advertisement services for both healthcare means providers and healthcare service providers. But this advertisement shouldn't invade any patient's privacy. It is important to be aware that advertisement is the most mature and trusted business model of the mobile internet ecosystem. Respect on such well-established business model is essential to initiate new businesses.

III. Ecosystem-Driven Standardization Efforts

Given the formulation of the new Health-IoT ecosystem, specific technical requirements can be derived more comprehensively and clearly. For example, the standardization of interfaces between any two actors within the ecosystem are necessary to ensure interoperability. The standardization of Health-IoT technologies should be ecosystem-driven instead of technology-driven. As shown in Fig. 3 three types of interfaces should be standardized.

A. *Interoperability of devices*

Firstly, the hardware and software interfaces between healthcare means suppliers and mobile application designers, and between means suppliers and mobile platform providers. For these interfaces, the Continua Health Alliance (CHA), a major standardization body working on device level interoperability, has recommended the Bluetooth Health Device Profiles (HDP), USB Personal Healthcare Device Profile and ZigBee Health Care Profile. They all apply a common data format specified by the ISO/IEEE 11073 family.

Based on these standards, the mobile platform providers and application designers can make a common driver for the same class of medical devices from different manufacturers. And then, the mobile devices can recognize a particular medical device according to its hardware descriptor and automatically apply correct data parsing and communication protocols. Thus, the complexity of patients' operation, hardware and software costs, and hence time to market, can be significantly reduced.

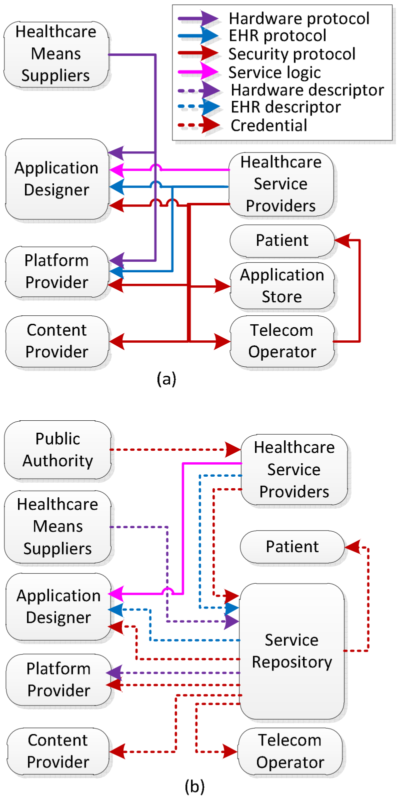


Fig. 3. Technical interfaces between actors (a) without standardization, and (b) with standardized hardware interfaces, data formats, and security schemes

B. *Electronic health record (EHR)*

Secondly, the format of electronic health record (EHR) should be standardized. The HL7, EN 13606 (specified by European Committee for Standardization), and ISO 18308 are the major efforts for this purpose. These EHR standards define 1) the protocol to exchange EHR messages; 2) the contents and structures EHR data, and 3) the mechanisms to ensure privacy and security of information sharing. By applying the EHR standards, the technical negotiations between healthcare service providers and mobile platform providers are simplified or hopefully avoided. The negotiation between healthcare service providers and application designers are simplified too.

C. *Uniform security mechanisms*

Thirdly, security schemes throughout the entire ecosystem should be standardized. Otherwise, all the parties would certainly intend to specify their own security mechanisms to protect their information as well as business benefits. However, the existing standardization efforts have not provided a solution so far. The root-cause is, the motivation of these standardization efforts on the security aspects is opposite to the "cooperative ecosystem". Device vendors and service providers mostly prefer to keep the security mechanisms to be proprietary aiming for the so-called "vendor lock-in". Unfortunately, it is the market itself that is "locked"!

In the solution that will be presented in the next section of this paper, to be trusted by the whole ecosystem, the traditional application store should be accredited by public authorities, and then it is transformed into a service repository. All security credentials are recorded by the repository and supervised by the public authority.

D. *Keep the service logic proprietary*

It is necessary to mention that the service logic should be left proprietary instead of standardized to encourage differential competition. The healthcare service providers can customize proprietary apps from application designers to accomplish specific value-added services. This point is supplementary to the efforts of IHE which promotes the coordinated use of established standards such as HL7 to address specific clinical need in support of optimal patient care.

IV. Ecosystem-Driven Security Mechanisms

A healthy ecosystem should protect the benefits of all stakeholders by balancing the control and avoiding monopoly. In the Health-IoT ecosystem, security mechanisms are the primary technical means to do so (of course, there are many non-technical measures that should be applied, but they are out of the scope of this paper). As the patients never believe the content providers and telecom operators can really protect their privacy [10], the only solution is to accredit an independent mediator, here the service repository, by the public authority. Principally, only the owner of the private information (the patient and his/her healthcare service provider) can access the information.

Based on the above considerations, the security schemes are proposed in Fig. 4 The public authority, service repository, healthcare service provider, content provider, telecom operator, and patient are the main actors in these security schemes.

A. *Public-based authentication*

As illustrated by the step1~14 to launch a particular Health-IoT service to the market, the enterprises should get authentication from the public authority first. The authentication is granted in the form of credential, so-called Secrete, which is a set of cryptography software running in the trusted hardware. The Secrete of each actor should be handed over by superior and safe approach. For example, it can be registered and delivered in person and delivered by accredited couriers. Only a certain person of a certain service provider can access a certain patient's information. That is, the authentication is tied to individuals instead of organizations.

B. *Repository-based credential management*

The service repository maintains all the Secretes of actors in trusted facilities. The content provider and telecom operator only maintain their own Secrets, which ensures the non-invasive message handover in-between patient and healthcare provider. The service provider maintains its own Secrete locally and can request other actors' keys from service repository.

C. *SE-based cryptography*

The secure element (SE) is a secured device that can store and execute specific cryptography algorithm. The algorithm is written by issuer and is extremely difficult to hack. A typical SE is the SIM (subscriber identification module) card that is

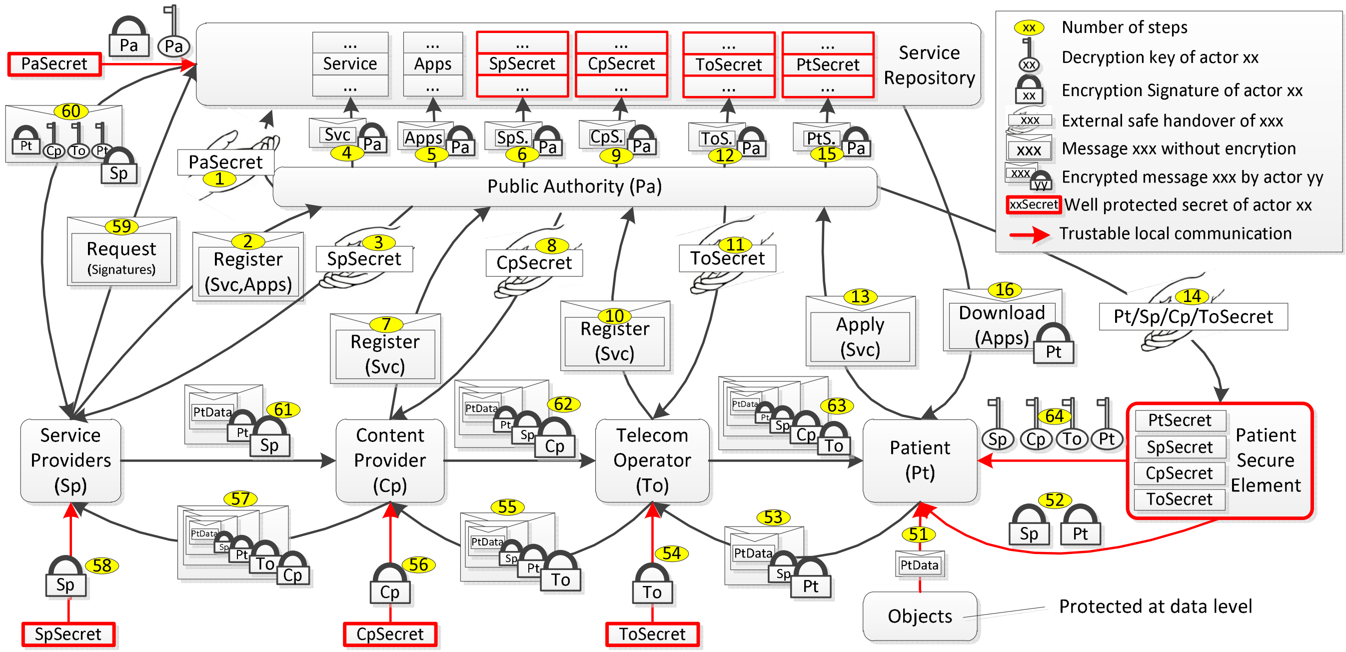


Fig. 4. The proposed security schemes

commonly used in mobile telecom services [12]. When a message is sent from the sender, it is encrypted and a signature is attached. Both the encryption key and signature are generated dynamically by the sender's SE. When the receiver receives the message, it sends the signature to its own SE. If the receiver has been authorized, there should be a copy (symmetric or asymmetric) of the sender's Secret in the receiver's SE, so that the receiver's SE can derive the decryption key. Only if everything (the signature, encryption key, decryption key, and Secrets in the SEs of both-sides) matches, the message can be decrypted. Furthermore, the Secrets in the SE are readable and writable only by the issuer (here the public authority). So, although the patient's SE contains the Secrets of other actors, the Secrets are not disclosed. And the communication between apps and SE is local within professional equipment (e.g. the Health-IoT Station). This reduces the security risk further.

If the SIM card is used as the SE, the logistics of SIM-card management will be significantly different with that in traditional telecom services. Traditionally, the SIM card is fully issued and supervised by the telecom operator. But in Health-IoT services, the SIM card should be issued and supervised by the public authority, and the telecom operator can only manage a part of the SIM as predefined by the public authority. This change may cause resistance by telecom operators. It should be resolved mainly by non-technical means, such as policy enforcement and financial compensation. And the telecom industry has also prepared technical solutions such as the remote subscription and SIM supervision [13].

D. *Non-invasive message handover*

The information from patient to service provider and response from service provider to patient are illustrated in the step 51~60 and step 61~64 respectively. As the content provider and telecom operator only maintain their own Secrets, they have no access to the messages transmitted through their facilities. In other words, the Health-IoT streams are transparent to them. This mechanism is called non-invasive message handover which is essential to get trusted from the end users and financial sources.

V. Ecosystem-driven Terminal Design

In order to establish the proposed Health-IoT ecosystem, the in-home healthcare station (IHHS) is the most important infrastructure that can be shared by all actors in the ecosystem. Therefore it should not be a close system. Instead, it should be open to third party applications like telecommunication and entertainment, so that e.g. the content suppliers and telecom operators can deliver other value-added services through it.

By installing specific apps, the IHHS can transform into many variants, from the logbook, to fatal monitor, wheel chain controller, portable monitor, smart walker, and medicine box (Fig.5). It is suitable for mass production with low cost as it is a "*standard*" product. It is also broadly acceptable by the whole ecosystem as it is based on an "*open*" platform.

Fig. 5. Hardware variants of the Health-IoT Station based on open platform

Before look into the technical details, several design principles for the terminal should be clarified first. These principles are the foundation for the design of hardware and software architectures of a particular solution. More details are also presented in a previous article [15].

A. *Common platform based on 3C products*

As required by the business ecosystem, the terminal hardware should be based on a mass produced 3C (computing, communication, consumer) platform such as mobile phone and tablet PC. The huge volume of 3C products has significantly minimized the development costs. The latest semiconductor and software techniques are adopted timely. So the IHHS providers can introduce their products with the shortest time-to-market and best performance-to-cost ratio. Performance, user experience, and maturity of developer community are the criteria for the platform selection. Well established software distribution channels are also important to establish the proposed Health-IoT ecosystem.

B. *Certification of the Health Extension*

The IHHS provider needs to customize the specific parts, so-called the Health Extension, of the terminal. These parts are not standard peripherals in the 3C platform, e.g. some specific mechanical structures and interfaces to biomedical sensors and device. The Health Extension is the importance entrance for the healthcare service providers and public authorities to be involved in the Health-IoT ecosystem. So, the IHHS provider must cooperate with them to offer a complete solution. In particular, the healthcare service providers can customize proprietary Health Extension based on their know-how and service contents. Only the certified products can access their contents. The public authorities can specify regulations to certify the Health Extension. Only the certified products and services can get the financial supports, such as the reimbursement of cost. This is an important technical measure to guarantee the Health-IoT is under public supervision.

C. *Interoperability and extendibility*

Devices from third parties are integrated to the IHHS through the device adaptors in the Health Extension. The aforementioned standardized interfaces for devices from various vendors should be supported, such as the ones recommended by the Continua Health Alliance: the Bluetooth Health Device Profiles, USB Personal Healthcare Device Profile and ZigBee Health Care Profile. However the IHHS developers cannot expect one single interface can fit all devices. Most of the popular interfaces should be supported such as USB, Bluetooth, WiFi, NFC (near field communication), RFID, WSN, etc. At the same time, extendibility for future standards should be reserved too.

D. *Convenient and trusted software distribution*

The application software for a particular Health-IoT service should be distributed through the application stores which have essentially enabled the mobile internet ecosystems. It is friendly both to the users and to the apps developers. The drivers for the Health Extension should be either natively integrated in the 3C platform or installed together with the apps, so that the device and service can be plug-and-play. At the same time, the publishing and installation of Health-IoT should be supervised by the public authorities. So the IHHS solution should provide trusted user identification, authentication and payment mechanisms e.g. through secured NFC interface. Solid security is ensured by the secure element which could be the SIM card, the NFC card, or an embedded secure device.

E. *Standardized and secured EHR handling*

The handling of EHR data is the center of information and service integration. For this purpose, a virtual local database is necessary to isolate the apps from different parties. The format of data exchanged between devices, apps, and backend systems should follow international standards such as the ISO/IEEE 11073, EN 13606, and ISO 18308 specifications. At the same time the access of EHR data should be authorized and controlled strictly, e.g. complying with the mechanisms proposed above.

F. *Efficient service composition*

The service oriented architecture (SOA) has been broadly adopted in the integration of healthcare services [15]. But to execute the intensive SOA middleware on the mobile platform is still challenging. State-of-the-art simplification and optimization are necessary.

G. *Efficient information integration*

The IHHS needs to efficiently process not only the vital signals from biomedical sensors but also the human activities collected by many other sensors such as microphone, camera and infrared sensor. The processing of multimedia data and diagnostic analysis of vital signals are computation-intensive. The popularity of multicore processors in 3C platforms has made it possible to perform such intensive tasks on a low cost IHHS [14]. Correspondingly, the terminal software architecture should be optimized to make the best use of the parallel computation capacity of multicore processors. This will be further introduced in the next section.

VI. ECOSYSTEM-DRIVEN DATA HANDLING

A. *Data flows*

As shown in Fig.6, the main data flows between the three parts of a Health-IoT system, the service backend, local station, and sensor devices are illustrated. The raw data are collected by the sensor devices with more or less compression because of the low processing performance of the sensor nodes. Here needs a trade-off between the processing speed, transmission power consumption, and price. Based on the data load distribution along the path, the nodes in the network are designed in two types called Main Nodes and Sub Nodes to maximize the utilization of the network resource. So the whole data flow starts from the sub nodes of the sensor network, path through the main nodes and be first packed and transmitted to the IHHS.

In the IHHS, local database is designed to manage the data flows and as the uniform data interface between the applications from various service suppliers. Usually the IHHS device is much more powerful than the sensor nodes, so higher efficiency data compression and visualization can be implemented. The IHHS should also receive data from the service backend e.g. the doctor side or the hospital side. These data could indicate the new schedule of medication, sampling data command, etc. These data are saved into the local database and presented to the user. At the same time compressed or re-packetized data are sent to the service backend.

More complex data analysis and diagnosis tools could be involved in the service backend. Bigger databases should be allocated to manage all the data from the whole system. Here service providers e.g. doctors analyze the uplink data and generate the feedback data. Feedback data will be sent back to the IHHS and presented to the user.

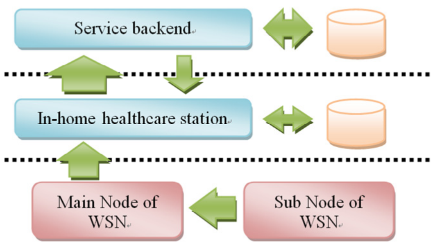


Fig. 6. Data flows in the Health-IoT system

B. Self-contained data formatting

As mentioned before, the standardized interfaces can improve the interoperability between devices and applications. Complementary approaches are also needed because 1) the number of standardized interfaces is still quite big; 2) it takes time until a few dominative standards are proven in practice; and 3) as the root cause, the ecosystem is open and welcomes new players and new solutions all the time.

As a complementary technical solution, a self-contained data formatting scheme is proposed. As shown in Fig.7, the data between the sensor devices and IHHS, and the IHHS between the service backend are packetized together with concrete descriptions. Each packet has its own header segment. At the beginning of the header segment, there is an optional title to identify the data packets e.g. from which device or patients. And the following bytes are the description bytes. It uses flags to describe the characters of the payload in a packet, which contains the data information.

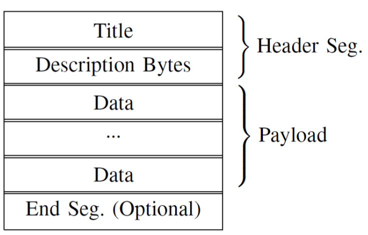


Fig. 7. Self-contained data formatting

The descriptions contain all the necessary information to pass and further process the data in the payload such as the packet size, how many samples are contained by this packet, which kind of sensor is used, the node id of the samples, compressing algorithm used for these samples, the compressing step, etc. For example, if the data is from a wearable ECG sensor device, its description segment may contain the manufacturer's information, sample rate, resolution, length, and the compressing/decompressing algorithms if the raw data is compressed.

Because the flag-scheme is used, the description parameters can be updated easily. New features can be integrated and unsupported features can be removed later. After the header segment, there is the data segment. It is the main body of the packet. It contains all the samples data. All the data should match the described structure. At the end of the packet there is an optional end segment. It indicated the end of the packet.

C. Layered compression

For most of data compression algorithms, after several compression phases, the data distortion caused by the compressing/decompressing usually gets worse and worse, and the computational complexity gets higher and higher. In the Health-IoT system, from the frontend sensor devices to the remote server, the data content is bigger and bigger, and the device processing speed is faster and faster. So the propose layered compression is to divide the whole data compression process into multiple steps. Based on different processing capacity of the devices, different process load is allocated. After each step, the data size becomes smaller and smaller. This will help to reduce the traffic load and transmission energy consumption of the second half of data path. Taking the advantage of the self-contained data formatting, data packets from one step contains all the information for the next step to decode and further compress the data.

For example, if the compression algorithm in the former step is based on a dictionary, this dictionary can be contained in the description of the packets. Moreover, because the system focuses on healthcare application and the data characteristics are stable, it is possible to use predefined parameters in the compressing algorithm. Saving these values together with the software in different devices, and then there is no need to send the compressing parameters with the data packets. Then along the whole data path, the packet format could keep the same.

VII. The Prototyping System

A. Application scenario

To verify the concepts and design strategy proposed in this paper, a prototype system has been implemented and evaluated in field trials. As a typical case of the IHHS, application scenario of an iMedBox system has been proposed in previous work [14-17]. The proposed IHHS solution is based on open source operation systems like the Google Android. It uses standard mobile internet terminal hardware, such as tablet PC, provided by various 3C (consumer, communication and computing) manufactures. The iMedBox system is a typical instance of the above terminal design principles.

As shown in Fig.8, a powerful intelligent medicine box (iMedBox) works not only as in-home medicine container, but also as a "*medication inspector*", and an "*onsite examiner*" in daily healthcare monitoring. It connects to the healthcare service providers through 3G and Wi-Fi networks and communicates to medical devices through USB, NFC, RFID, and WSN (wireless sensor network). Medication and fatal information is transmitted to the backbone systems and feedback from service providers is presented to the patient at home.

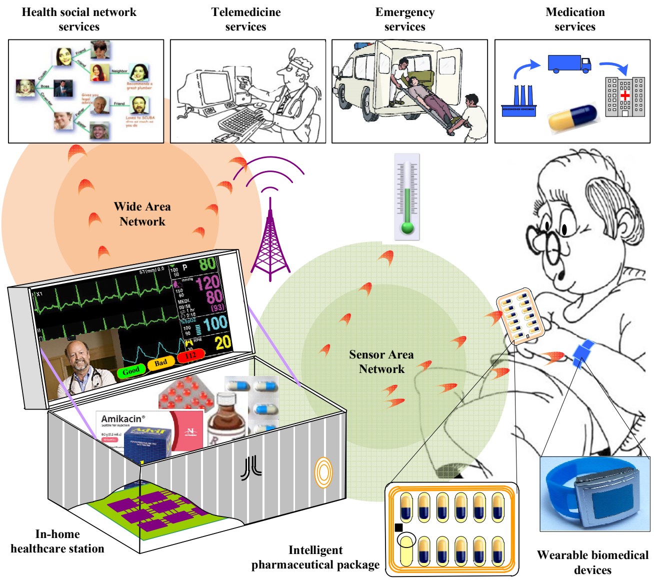


Fig. 8. The application scenario of the In-Home Healthcare Station

B. *The iMedBox prototyping system*

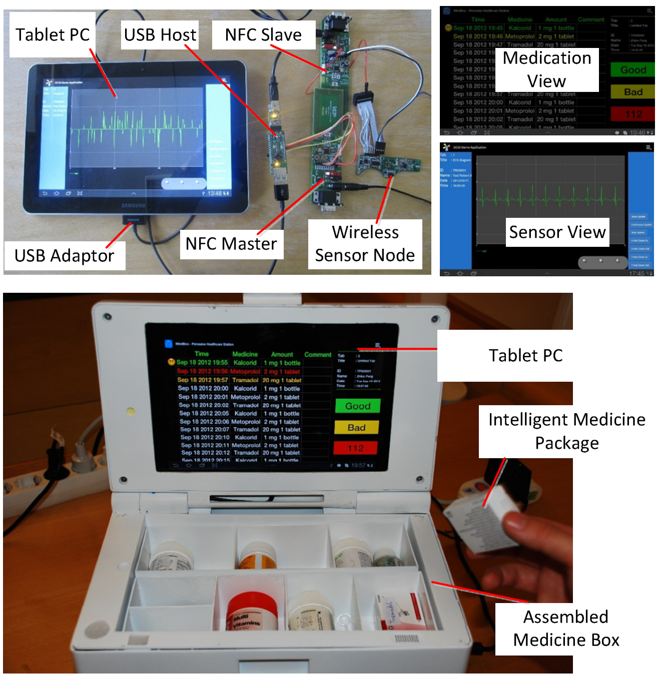


Fig. 9. Implementation of iMedBox prototype

A prototyping system of the iMedBox is implemented based on the Samsung Galaxy Tab10.1 tablet PC. It has a 10.1 inch display with touch screen, a dual-core 1GHz ARM Cortex-A9 CPU, and connections through USB OTG, Wi-Fi and 3G. The operation system is Android 3.1. As shown in Fig.9, we extended the hardware through a USB adaptor to support NFC and WSN connections which are not standard peripherals of tablet PC so far. A functional iMedBox is assembled by embedding the tablet and extension modules into a hand-molding box. As a part of the demonstration, intelligent medicine packages are made by attaching inlay RFID tags onto ordinary medicine packages.

The application software for the demonstration is implemented in Java as standard Android apps. The graphic engine is the AChartEngine, and database engine is the SQLite. A dedicated data processing engine for data packet parsing, a security engine for authentication and cryptograph, and a web server for 3G/WiFi connection are implemented based on the basic Android 2.1 API which is supported by major variants of Android. Two types of GUI are designed so far: a Sensor View for sensor data and a Medication View for prescription.

C. *Users' feedback and improvement directions*

Some field trials have been carried out in nursing centers and elderly houses in Blekinge, Sweden. The system concepts have been confirmed by the positive feedback. The medication reminding and recording functions can significantly improve the medication compliance especially for elderly. Seamless integration to the hospital's prescription system is necessary to reduce the workload of manual input. The proposed authentication scheme sounds complicated but necessary to reassure the users.

Some insufficiencies are also pointed out. The user interfaces are still too complicated for elderlies although they are acceptable for nurses. The texts are not clear enough. This can be improved by replacing the Android's default colors and fonts which is quite "fashion" but lack of clarity. The network connection and authentication are too slow (currently around 10 seconds). We will further measure the latency step-by-step to find out the bottleneck and improve.

VIII. Conclusion and Future Work

To develop successful Health-IoT solutions towards the IoT, a cooperative ecosystem should be established first. Technical architectures should be centered to the ecosystem especially regarding the interoperability, security and information system integration.

In this paper, a cooperative ecosystem of Health-IoT is formulated based on the reconstruction of the traditional healthcare and mobile internet ecosystems. To support this ecosystem, the new ecosystem-driven strategies are applied in all aspects of the development of a Health-IoT solution. As show cases, the design principles or guidelines for standardization efforts, security mechanisms, open-platform-based terminals, and data handling schemes are presented with details. To support fair distribution of benefits among all stakeholders, value-centric security schemes are proposed, including the public authority-based authentication, the secure element (SE) based cryptography, and the non-invasive message handover. In order to achieve the economy of scale, an open-platform-based IHHS solution is proposed as a universal platform for device and service integration and convergence. Several important design principles are given including common platform based on 3C products, certification of the Health Extension, interoperability and extendibility, convenient and trusted software distribution, standardized and secured EHR handling, efficient service composition, and efficient information integration. Corresponding data handling schemes are designed including the layered data compression and self-contained data formatting.

Finally, to verify the concepts and technical feasibilities, we have developed a prototype system called iMedBox. It is a specific case for medication management and in-home monitoring applications. The iMedBox hardware, software and backbone system are implemented and evaluated by field demonstrations. The positive feedbacks have proven the feasibility of proposed design methods, proposed architectures and solutions. Based on the results of this paper, economically feasible services are closer to reality.

One important future work is to implement the proposed security mechanisms by involving more external partners including the SIM card maker and telecom operator. Then we plan to integrate the iMedBox terminal into some existing healthcare information systems and carry out more trials.

REFERENCES

[1] Daniele Miorandi, et al. "Internet of things: Vision, applications and research challenges", *Ad Hoc Networks*, online 21 April 2012

[2] Mari Carmen Domingo, "An overview of the Internet of Things for people with disabilities", *J. Network and Computer Applications*, 35(2), March 2012, 584-596

[3] Dohr, A.; et al. "The Internet of Things for Ambient Assisted Living", *Int. conf. information Technology: New Generations*, 2010, 804 – 809

[4] Hande Alemdar, Cem Ersoy, "Wireless sensor networks for healthcare: A survey Original", *Computer Networks*, 54(15), 2010, 2688-2710

[5] Chang Liu, et al. "Status and trends of mobile-health applications for iOS devices: A developer's perspective" *J. Systems and Software*, 84(11), November 2011, 2022-2033

[6] Predrag Klasnja, Wanda Pratt, "Healthcare in the pocket: Mapping the space of mobile-phone health interventions", *J. Biomedical Informatics,* 45(1), 2012, 184-198

[7] Inmaculada Plaza, et al. "Mobile applications in an aging society: Status and trends", *J. Systems and Software,* 84(11), 2011,1977-1988

[8] Google Android operation system, http://www.android.com

[9] Google Health, www.google.com/health, accessed on May 12, 2012.

[10] Brian Dolan, "10 Reasons why Google Health failed", *MobiHealthnews by Chester Street Publishing, Inc,* Jun 27, 2011

[11] ITU, "The Internet of Things-Executive Summary", www.itu.int, 2005

[12] Kalman, G.; Noll, J.,"SIM as Secure Key Storage in Communication Networks", *Int. conf. ICWMC*, 2007, 55 – 55

[13] Luis Barriga, et al. "M2M Remote-Subscription Management", *Ericsson Review* 2011 (1)

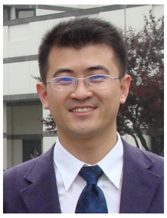
[14] Zhibo Pang, Qiang Chen, Lirong Zheng, "A Pervasive and Preventive Healthcare Solution for Medication Noncompliance and Daily Monitoring", *2nd Inte. Symp. on Applied Sciences in Biomedical and Communication Technologies (ISABEL 2009)*, pp1-6, Nov. 2009

[15] Zhibo Pang, Lirong Zheng, Junzhe Tian, Sharon Kao-Walter, Elena Dubrova , Qiang Chen. "Design of a Terminal Solution for Integration of In-home Healthcare Devices and Services towards the Internet-of-Things", *Enterprise Information Systems*, DOI:10.1080/17517575.2013.776118, April 2013.

[16] Zhibo Pang, Qiang Chen; Junzhe Tian, Lirong Zheng, Elena Dubrova. "Ecosystem Analysis in the Design of Open Platform-based In-Home Healthcare Terminals towards the Internet-of-Things". *International Conference on Advanced Communications Technology (ICACT).* Jan 2013, Pyeongchang, Korea..

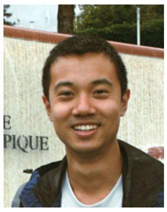
[17] Zhibo Pang, Qiang Chen; Lirong Zheng, Elena Dubrova. "An In-home Medication Management Solution Based on Intelligent Packaging and Ubiquitous Sensing". *International Conference on Advanced Communications Technology (ICACT)*. Jan 2013, Pyeongchang, Korea.

[18] Zhibo Pang, "Technologies and Architectures of the Internet-of-Things (IoT) for Health and Well-being", PhD Thesis, Royal Institute of Technology (KTH), Stockholm, Sweden, 2013.



Zhibo Pang received B.Eng. degree in Electronic Engineering from Zhejiang University, Hangzhou, China, in 2002, MBA in Innovation and Growth from University of Turku, Turku, Finland, in 2012, and PhD in Electronic and Computer Systems from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 2013.

He is a research scientist at ABB Corporate Research, Västerås, Sweden. Before joined ABB, he worked as technical manager in semiconductor industry, designing baseband and application processors and turn-key solutions for mobile smart devices. He has 15 patents and over 30 peer-reviewed papers in international journals and conferences. He was awarded the National Great Invention Award by the Ministry of Information Industry of China in 2005, won the First Place Prize of the RFID Nordic EXPO in 2008 and Outstanding Paper Awards in ICACT2013. His current research interests include the Internet-of-Things, wireless sensor network, industrial communication, real time embedded system, enterprise information systems, automation networks, and multicore system-on-chip and network-on-chip. He also works on the business-technology joint research such as business model design, value chain formulation, strategy, and entrepreneurship & intrapreneurship.



Junzhe Tian received B.Eng. degree in Communication Engineering from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2010, M.Sc. degree in System-on-Chip Design from Royal Institute of Technology (KTH), Stockholm, Sweden, in 2013. After graduate, he became a Software developer at Excosoft AB Stockholm, Sweden. Currently he is working on design and implementation of Graphical Editing Framework (GEF) and Document Type Definition (DTD) in publishing industry.

Concept-Based Readability Measurement and Adjustment for Web Services Descriptions

Pananya Sripairojthikoon, Twittie Senivongse

Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand

pananya.s@student.chula.ac.th, twittie.s@chula.ac.th

***Abstract*—Web Services is a technology for building distributed software applications that are built upon a set of information and communication standards. Among those standards is the Web Services Description Language (WSDL) which is an XML-based language for describing service descriptions. Service providers will publish WSDL documents of their Web services so that service consumers can learn about service capability and how to interface with the services. Since WSDL documents are the primary source of service information, readability of WSDL documents is of concern to service providers, i.e., service descriptions should be understood with ease by service consumers. Providing highly readable service descriptions can then be used as a strategy to attract service consumers. However, given highly readable information in the WSDL documents, competitors are able to learn know-how and can copy the design to offer competing services. Security attacks such as information espionage, client impersonation, command injection, and denial of service are also possible since attackers can learn about exchanged data and invocation patterns from WSDL documents. While readability of service descriptions makes Web services discoverable, it contributes to service vulnerability too. Service designers therefore should consider this trade-off when designing service descriptions. Currently there is no readability measurement for WSDL documents. We propose an approach to such measurement so that service designers can determine if readability is too low or too high with regard to service discoverability, service imitation, and service attack issues, and then can consider increasing or lowering service description readability accordingly. Our readability measurement is based on the concepts or terms in service domain knowledge. Given a WSDL document as a service description, readability is defined in terms of the use of difficult words in the description and the use of words that are key concepts in the service domain. As an example, we measure readability of the WSDL document of E-commerce Web services, and experiment on redesigning of WSDL terms to adjust readability.**

***Keyword*—Concept Hierarchy, Ontology, Readability, Web Services**

Manuscript received August 28, 2013.

P. Sripairojthikoon is with the Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok 10330, Thailand. (e-mail: pananya.s@student.chula.ac.th).

T. Senivongse is with the Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok 10330, Thailand. (Corresponding author phone: +66 2 2186996; fax: +66 2 2186955; email: twittie.s@chula.ac.th).

I. INTRODUCTION

WEB services is a technology for building distributed software applications. The building blocks are services which are software units that are built upon a set of information and communication standards. Among those standards is the Web Services Description Language (WSDL) which is an XML-based language for describing service descriptions. A WSDL document is defined by a service provider and used by service consumers in discovering service capability and establishing interaction between consumer-side applications and the Web service. The structure of a WSDL document (version 1.1) which describes what the service is capable of and what data are exchanged comprises the XML elements <types>, <message>, <portType>, and <documentation> [1].

Since WSDL documents are the primary source of service information, readability [2] of WSDL documents or Web services descriptions is of concern to service providers, i.e., service descriptions should be understood with ease by service consumers. Meaningful names should be given to the service interface, operations, input and output messages, and data. In addition, sufficient documentation should be provided regarding functional scope and limitation of use. Providing well-defined readable service descriptions can be used as a strategy by service providing organizations to attract service consumers.

Despite being desirable, readability of service descriptions has its downside. Other organizations can gather information from a service WSDL document to learn know-how and then copy the design to offer competing services. Also, published WSDL documents can provide security attackers with information like schemas of exchanged data, invocation patterns, and service location. Attackers may be able to guess other private operations. This leads to more serious attacks such as information espionage, client impersonation, command injection, and denial of service.

While readability of service descriptions makes Web services discoverable, it contributes to service vulnerability too. Service designers therefore should consider this trade-off when designing service descriptions. Currently there is no readability measurement for WSDL documents. We therefore aim to propose an approach to such measurement so that service designers can determine if readability is too low or too high with regard to service discoverability, service imitation, and service attack issues, and then can consider increasing or lowering service description readability accordingly.

We apply the concept-based readability measurement model proposed by Yan et al. [3] to the context of Web service descriptions. As the name implies, our readability measurement is based on the concepts or terms in service domain knowledge. Given a WSDL document as a service description, readability is defined in terms of the use of difficult words in the description and the use of words that are key concepts in the service domain. For example, if the service description contains simple words or closely-related terms within the domain, it should be easy to understand the functionality of the service from the service description. Here, service domain knowledge is described as ontology [4] that defines vocabulary of concepts and properties as well as their relationships, using an XML-based OWL language [5]. Readability assessment can be conducted by the quality assurance team or service designers who have knowledge of the service domain, and can be of several uses. The assessors can compare readability of their WSDL documents with that of the competing services and may evaluate if readability should be improved to attract more consumers. In a certain case, the assessors may consider adjusting the service descriptions if security issues are of concern. We also outline a method to increase and lower readability of service descriptions, and experiment on redesigning of WSDL terms of E-commerce Web services to adjust readability.

The rest of this paper is organized as follows. Section II discusses related work and Section III presents the concept-based document readability model. We propose a methodology to assess readability of WSDL documents and how to adjust it in Section IV. In Section V, we present an experiment on readability measurement for three Web services in the E-commerce domain for comparison, followed by an experiment to adjust readability of their WSDLs in Section VI. Section VII concludes the paper with a discussion of the approach and future outlook.

II. Related Work

By definition, readability means "the level of ease or difficulty with which text material can be understood by a particular reader who is reading that text for a specific purpose" [2]. Readability is dependent upon many characteristics of both the text and the readers, and its concept has been applied to many kinds of text material including books, technical documents, online documents, and Web pages. Many formulas for measuring text readability are available and most of them deal with only text features. That is, texts that use difficult words are more difficult to understand than those with simple words, and texts with long sentences and complex syntax are difficult to read.

Yan et al. [3] propose a different but interesting approach to measuring text readability in the context of online documents. They argue that not only domain experts but also average users are searching more and more for domain-specific information from online documents, particularly in the medical area, and these documents are of different readability level. However, traditional readability formulas are designed for general purpose texts and insufficient to deal with technical materials in a specific domain. Therefore, for the document ranking purpose, their model takes advantages of a traditional readability formula and domain knowledge to measure readability of domain-specific documents at the word level, i.e., it focuses on how the domain-specific terms in a document affect readability of the document. For example, if the document contains closely-related terms in the domain vocabulary, it should be more easily readable and comprehensible to readers of that domain. In their approach, the domain knowledge is represented as a concept hierarchy or ontology.

Zhao and Kan [6] argue that the ontology-based approach such as [3] has a limitation in that it requires expert knowledge which is still expensive and not readily available in most domains. They hence present an iterative computation on a resource-concept graph based on the intuition that readability of the domain-specific resources (or documents) and difficulty of domain-specific concepts provide accurate estimations of each other. The algorithm constructs a graph that represents what concepts are contained in a particular resource and what resources contain a particular concept. Then readability computation for the resources is based on a simple mutually recursive observation as follows: (1) a domain-specific resource *A* is less readable than another domain-specific resource *B* if *A* contains more difficult domain-specific concepts than *B*, and (2) a domain-specific concept *A* is more difficult than another domain-specific concept *B* if *A* is mentioned in less readable domain-specific resources than *B*. This approach is effective and less domain-dependent but it requires multiple resources in order to determine their readability, i.e., readability of any single resource cannot be determined on its own.

The research by Jatowt et al. [7] attempts to measure readability of Web pages based on link structure. It is motivated by their study on the correlation between readability of the source pages and that of the linked pages, i.e., there is a trend that Web pages would link to other pages with generally the same level of difficulty. They are also inspired by the TrustRank algorithm in which scores are propagated from good pages in an attempt to separate useful Web pages from spam. Therefore, their algorithm utilizes the Web link structure by propagating the readability scores of the source pages to the linked pages. This approach is useful since it becomes possible to measure the readability of Web pages that have little texts and to complement traditional readability measures which rely only on textual content.

It is seen that several research attempts propose different ways to determine the readability score of a document but none of them address measurement of WSDL documents readability. Since a WSDL document is self-contained and we should be able to determine its readability using its own content, we will adopt the concept hierarchy-based model by Yan et al. [3]. This model applies at the word level, meaning that it is based on concepts or terms within a document, rather than sentences or paragraphs; this suits well with the case of a WSDL document. Using this model, we then rely on the availability of the service domain ontology on the Internet. In the case that no domain ontology is available, we argue that an approach such as [8], which builds domain ontology using terms from the WordNet database [9], can be taken.

III. Concept-Based Document Readability Model

The concept-based readability model considers difficulty of terms in a WSDL document and two features of the WSDL document, namely document scope and document cohesion, with respect to the presence of domain terms in the WSDL document. As mentioned earlier, vocabularies of the service domain is formed as a concept hierarchy. The terms in the document which have a match in the concept hierarchy are regarded as domain terms; otherwise they are non-domain terms. The detail of the model components [3] is as follows.

A. Document Scope (DS)

Document scope is defined as the coverage of the domain concepts in the document. The coverage is viewed from two angles. First, the more the document contains domain terms, the less readable the document tends to be since the document is likely to contain a larger number of specific concepts. Second, the deeper the domain terms appear in the concept hierarchy, the more difficult the document is to read. The document scope *DS* of a WSDL document d_i can be computed by (1):

$$DS(d_i) = e^{-(\sum_{i=1}^{n} depth(c_i))} \quad (1)$$

where $depth(c_i)$ = depth of domain concept c_i in the WSDL document d_i, with regard to the concept hierarchy.

B. Document Cohesion (DC)

Document cohesion refers to how focused the text is on a particular topic. It can be computed by the semantic relatedness between the domain terms in the document which is reflected by the links (or shortest path) between them with respect to the given concept hierarchy. The more cohesive the domain terms in the document are, the more readable the document is. The document cohesion *DC* of a WSDL document d_i can be computed by (2)-(4):

$$DC(d_i) = \frac{\sum_{i,j=1}^{n} Sim(c_i, c_j)}{Number\ of\ Associations}, where\ n > 1, i < j \quad (2)$$

$$Sim(c_i, c_j) = -log \frac{len(c_i, c_j)}{2D} \quad (3)$$

$$Number\ of\ Associations = \frac{n(n-1)}{2} \quad (4)$$

where $len(c_i,c_j)$ = shortest path between c_i and c_j in the concept hierarchy,

D = maximum tree depth in the concept hierarchy, and

n = total number of domain concepts in the WSDL document d_i.

C. Simplified Dale-Chall's Readability Index (DaCw)

A well-known traditional readability formula is the Dale-Chall's readability index [10]. This index sees that, the length of the sentences in a document and the difficulty of words correlate with the difficulty of reading material. Since the concept-based readability model measures readability at the word level, sentence-level complexity is not applicable and hence only word difficulty is considered. To determine the difficulty of words, words in the document are identified as either familiar or unfamiliar words. That is, they are familiar words if they can be found in the Dale list of approximately 3,000 familiar words. Otherwise, they are unfamiliar, and hence difficult, words. The simplified version of the Dale-Chall's readability index *DaCw* of a WSDL document d_i can be computed by (5):

$$DaCw(d_i) = PDW \quad (5)$$

where *PDW* = percentage of difficult words in the WSDL document d_i (i.e., number of difficult words divided by number of words and multiplied by 100).

We will use the word-level model above to measure readability of WSDL documents since contents of WSDL elements are mainly words, and not sentences.

D. Concept-Based Readability

The concept-based readability model which determines the overall readability score of a WSDL document d_i can be computed by (6):

$$CRS(d_i) = DS(d_i) + DC(d_i) + DaCw(d_i)^{-1} \quad (6)$$

IV. Concept-Based Readability Assessment Methodology

The concept-based readability of a WSDL document can be measured by an assessor who is a member of the quality assurance team or a service designer. The assessor must have knowledge of the service domain in order to choose (or construct) the concept hierarchy or domain ontology appropriately, and at the end evaluate the readability score. The assessment methodology comprises the following steps.

A. Service Information Preparation

To prepare for assessment, the assessor first does the following.

Select WSDL Document

The assessor selects a Web service and acquires its WSDL document. Note that if the assessor wants to compare readability of two Web services, both services will be assessed individually but they must be in the same domain and share the same concept hierarchy.

Obtain Concept Hierarchy of Service Domain

The concept hierarchy can be a domain ontology defined by domain experts. We consult either search engines or Web sites that publish domain-specific ontologies, e.g., [11]-[13], to discover relevant ontology for the Web service. In the case that multiple ontologies of the service domain are found, a tool like Protégé [14] can be used by the assessor to merge them into a single integrated ontology.

If relevant domain ontology is not available for the Web service, WordNet can be used to help generate the concept hierarchy [8]. The key terms in the WSDL document will be extracted and interwoven with their hypernym terms from WordNet to build up the concept hierarchy. Nevertheless, obtaining the concept hierarchy through ontology libraries is preferred to obtaining it through concept hierarchy generation. This is because concepts in domain ontology are defined by domain experts and can include specific terms of the domain which better reflect domain knowledge, whereas concepts taken from WordNet are likely to be more generic terms. Only when the Web services are built for specific domains and appropriate domain ontology is not readily available should the concept hierarchy be generated from WordNet.

B. Readability Measurement

Generally the assessor measures readability of the whole WSDL document, but in some case it might be useful to measure readability of particular WSDL elements for a more detailed analysis. The assessor can first select to measure readability of either the whole document or certain elements, and then apply the concept-based readability model (6). In the case that the measurement targets a particular WSDL element, the content of the element corresponds to a textual document of the model.

To support WSDL readability measurement, we develop a tool called WSDL Readability Calculator which is shown in Fig. 1. The tool is developed by using Eclipse Java EE IDE [15]. Protégé Java library [14] is used to read the domain ontology file. The tool requires the following input from the assessor:

- Web service description URL or a WSDL file;
- Service domain ontology URL or an OWL file;
- Calculation method, i.e., whole document or specific WSDL element.

Once the assessor inputs all service details and starts the calculation, the tool will perform the seven steps below:

Step 1: Extract terms from WSDL elements. WSDL syntax (such as tag names, types, cardinality) is excluded; only their contents will be considered. Duplicate terms are also removed. The following is an example of an input message element of a WSDL operation and a term extracted from this element.

- Input message: <wsdl:input message="services:CreateInvoiceRequest" name="CreateInvoiceRequest" />

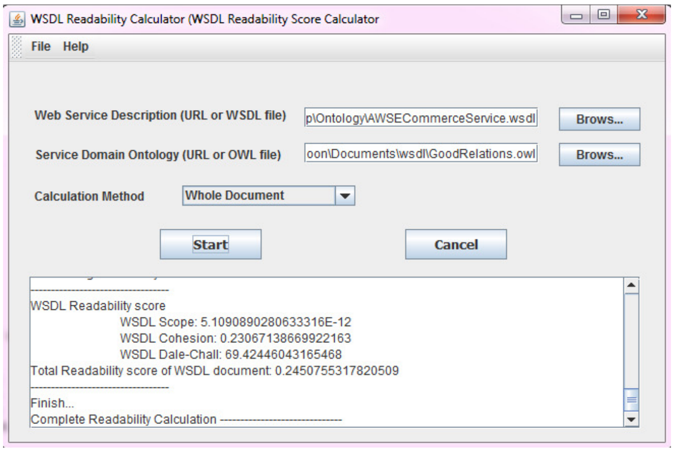


Fig. 1. WSDL readability calculator.

- Extracted term: CreateInvoiceRequest

Step 2: Extract key individual terms from the list of extracted terms in step 1. The terms will be changed to their singular form. Duplicate terms and single character will be removed. For example, the term "CartItems" will be extracted into two individual terms: "Cart" and "Item".

Step 3: Map key individual terms to the concepts in the concept hierarchy. To measure readability score of a WSDL document, only the WSDL terms that are present in the concept hierarchy will be used for measuring the document scope and document cohesion. On the other hand, all terms in a WSDL document will be used for measuring the simplified Dale-Chall's readability index.

Step 4: Calculate the document scope using (1). To calculate the document scope, the depth of each domain concept found in the WSDL document is determined with regard to the concept hierarchy.

Step 5: Calculate the document cohesion using (2)-(4). To calculate the document cohesion, the shortest path between each pair of domain concepts that are found in the WSDL document is determined, with respect to the concept hierarchy.

Step 6: Calculate the simplified Dale-Chall's readability index using (5). To calculate the simplified Dale-Chall's readability index, the percentage of difficult words not found in the Dale list of familiar words is determined.

Step 7: Calculate the concept-based readability score using (6). The readability score of a WSDL document is calculated using the concept-based readability model.

C. Evaluation

Once the readability score is obtained, the assessor evaluates if readability of the whole WSDL document (or a particular WSDL element) is appropriate. The assessor may be concerned about attracting the consumers and at the same time being vulnerable to attackers and competitors. The assessor can use the readability score for comparison purpose, e.g., comparing with the scores of competitors' services. The scores can be adjusted if the assessor sees fit.

D. *Readability Adjustment*

The concept-based readability score is dependent on the design and naming of terms within the WSDL document and the quality of the concept hierarchy. That is, the assessor should choose service domain ontology from a reliable source.

Measuring readability of specific WSDL elements can help the assessor to pinpoint which parts of the WSDL document should be redesigned and how to adjust their contents in order to increase or lower the score appropriately. According to (6), it is apparent that we can adjust the readability score by redesigning or renaming of terms within the WSDL document since change of terms can affect the document scope, document cohesion, and simplified Dale-Chall's readability index in the following ways.

- Document Scope: For a WSDL document, adding terms or changing to deeper terms in the concept hierarchy will decrease its readability score.
- Document Cohesion: For a WSDL document, adding terms or changing to terms that are more cohesive, i.e., more closely associated, with respect to the concept hierarchy will decrease the shortest path and hence increase its readability score.
- Simplified Dale-Chall's readability index: For a WSDL document, adding or changing to terms in the Dale list of familiar words will increase its readability score.

Therefore, to increase the readability score to improve service discoverability, the assessor may, where appropriate,

- Change domain terms in the WSDL document to non-domain terms;
- Change domain terms in the WSDL document to those that are shallower in the concept hierarchy;
- Change or add domain terms so that the WSDL document contains domain terms that are more cohesive with respect to the concept hierarchy;
- Change or add terms so that the WSDL document contains more of Dale's familiar words.

On the other hand, to lower the readability score to make the service description less comprehensible to competitors and attackers, the assessor may, where appropriate,

- Change non-domain terms in the WSDL document to domain terms;
- Change domain terms in the WSDL document to those that are deeper in the concept hierarchy;
- Change or add domain terms so that the WSDL document contains domain terms that are less cohesive with respect to the concept hierarchy;
- Change or add terms so that the WSDL document contains less of Dale's familiar words.

After redesigning or renaming of terms within the WSDL document, the assessor will repeat the measurement process to obtain the adjusted score. The adjustment and measurement can be repeated as necessary until the assessor is satisfied with the score.

V. EXPERIMENT ON READABILITY MEASUREMENT

As an experiment, this section describes the application of the methodology in Section IV to three Web services in the E-commerce domain, i.e., Amazon [16], PayPal [17], and eBay [18]. The WSDLs of these services are processed by the WSDL readability calculator to determine readability of the whole WSDL documents. Note that, to measure readability of a specific type of WSDL elements, the methodology can be applied in a similar manner.

For demonstration purpose, we show step-by-step application of the methodology to measure readability of the whole WSDL of Amazon Web service.

A. *Service Information Preparation*

Amazon Web service delivers a set of E-commerce services, i.e., payment management, stock management, and package and shipping management. Its WSDL can be found at [16]. The GoodRelations ontology [19] is selected as the concept hierarchy for the E-commerce domain. Part of it is depicted in Fig. 2. A number of concepts such as account, contact information, payment method, etc. are listed in this ontology.

B. *Readability Measurement*

The process to measure readability of Amazon's WSDL is as follows.

Step 1: Extract Terms from WSDL Elements

Element names are extracted from WSDL syntax. Duplicate terms are also removed. In total, there are 373 extracted terms. Table I shows some of the terms that are extracted from a WSDL snippet in Fig. 3.

Step 2: Extract Key Individual Terms

The extracted terms are further extracted to obtain key individual terms and are changed to a singular form. Again, duplicate terms and any single character are removed. For example, the term "SimilarProducts" will be extracted into two individual words "Similar" and "Product". In total, there are 275 individual terms, and Table II shows some of them.

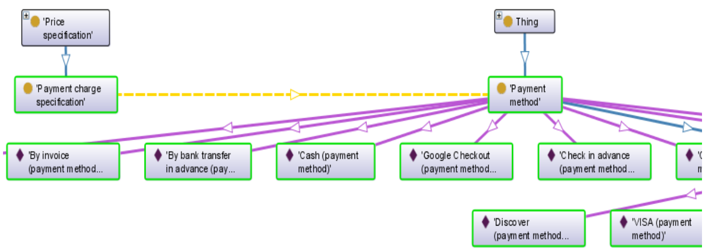


Fig. 2. Part of GoodRelations ontology.

TABLE I
EXAMPLE OF EXTRACTED TERMS FROM AMAZON'S WSDL

PurchaseURL	SimilarProducts
MobileCartURL	TopSellers
SubTotal	TradeInValue
CartItems	SimilarViewedProducts
SavedForLaterItems	LowestNewPrice

```
<xs:element ref="tns:Request" minOccurs="0" />
<xs:element name="CartId" type="xs:string" />
<xs:element name="HMAC" type="xs:string" />
<xs:element name="URLEncodedHMAC" type="xs:string" />
<xs:element name="PurchaseURL" type="xs:string" minOccurs="0" />
<xs:element name="MobileCartURL" type="xs:string" minOccurs="0" />
<xs:element name="SubTotal" type="tns:Price" minOccurs="0" />
<xs:element ref="tns:CartItems" minOccurs="0" />
<xs:element ref="tns:SavedForLaterItems" minOccurs="0" />
<xs:element ref="tns:SimilarProducts" minOccurs="0" />
<xs:element ref="tns:TopSellers" minOccurs="0" />
<xs:element ref="tns:NewReleases" minOccurs="0" />
<xs:element ref="tns:SimilarViewedProducts" minOccurs="0" />
<xs:element ref="tns:OtherCategoriesSimilarProducts" minOccurs="0" />
...
<xs:element name="TrackSequence" type="xs:string" minOccurs="0" />
<xs:element name="TradeInValue" type="tns:Price" minOccurs="0" />
<xs:element name="UPC" type="xs:string" minOccurs="0" />
<xs:element name="LowestNewPrice" type="tns:Price" minOccurs="0" />
<xs:element name="LowestUsedPrice" type="tns:Price" minOccurs="0" />
```

Fig. 3. Snippet of Amazon's WSDL.

TABLE II
EXAMPLE OF INDIVIDUAL TERMS FROM AMAZON'S WSDL

Purchase	Url	Mobile	Cart
Sub	Total	New	Price
Saved	For	Later	Viewed
Similar	Product	Top	Seller
Item	Trade	In	Value

Step 3: Map Key Individual Terms to Concepts in Concept Hierarchy

Individual terms will be mapped to the concepts in the GoodRelations concept hierarchy. In total, there are 21 individual terms that can find a match in the concept hierarchy. Table III shows some of these domain terms.

Step 4: Calculate Document Scope

The depth of each domain term found in step 3 is determined with regard to the concept hierarchy. Table IV shows the depth of some domain terms. Given (1), the document scope of Amazon's WSDL document is 5.1091E-12.

Step 5: Calculate Document Cohesion

The shortest path between each pair of 21 domain terms is determined with respect to the concept hierarchy of depth 5. Table V shows the shortest path between some of them. Given (2)-(4), the document cohesion of Amazon's WSDL document is 0.2307.

TABLE III
EXAMPLE OF DOMAIN TERMS FOUND IN AMAZON'S WSDL

Item	Product
Value	Price

TABLE IV
EXAMPLE OF DEPTH OF DOMAIN TERMS IN AMAZON'S WSDL

Concept c_i	Depth
Item	3
Product	2
Value	1
Price	1

TABLE V
EXAMPLE OF SHORTEST PATHS BETWEEN PAIRS OF DOMAIN TERMS IN AMAZON'S WSDL

c_i	c_j	$len(c_i, c_j)$
Item	Product	6
Item	Value	6
Item	Price	5
Product	Value	5
Product	Price	4

Step 6: Calculate Simplified Dale-Chall's Readability Index

There are 86 out of 275 individual terms which are found in the Dale list of familiar words. Table VI shows some of them. Therefore there are 189 terms that are considered difficult. Given (5), the simplified Dale-Chall's readability index of Amazon's WSDL document is 68.7273.

Step 7: Calculate Concept-Based Readability Score

Given (6), the concept-based readability score of Amazon's WSDL document is 0.2452.

C. *Readability Comparison*

Similarly to Amazon's WSDL, the WSDLs of PayPal and eBay can be processed to determine readability based on GoodRelations E-commerce ontology. Table VII shows the measurements of all three services in comparison. PayPal's Web service provides the most readable WSDL document whereas readability of Amazon's WSDL is the lowest. Readable and comprehensible service description can be one factor in the popularity of PayPal Web service, while WSDL readability of eBay, and Amazon follows PayPal's very closely, indicating that they can be competing Web services providing similarly readable service descriptions. On the other hand, Amazon's WSDL document is likely to expose less information to competitors and attackers.

VI. EXPERIMENT ON READABILITY ADJUSTMENT

After the readability score is obtained, the assessor evaluates if readability of the whole WSDL document (or a particular WSDL element) is appropriate. If not, the assessor determines whether readability score of WSDL document should be increased or lowered compared to the current score.

TABLE VI
EXAMPLE OF FAMILIAR WORDS FOUND IN AMAZON'S WSDL

For	Cart	Top
New	Value	Price

TABLE VII
READABILITY MEASUREMENTS

	Amazon	PayPal	eBay
Document Scope	5.1091E-12	1.0262E-10	5.7495E-19
Document Cohesion	0.2307	0.2989	0.2875
Simplified Dale-Chall's Readability Index	68.7273	59.659	79.5482
Concept-Based Readability Score	0.2452	0.3156	0.3001

Adjustment of WSDL readability score can be done according to the criteria that have been described in Section IV.D. With the adjustment process, readability score calculated from the document scope, document cohesion, and simplified Dale-Chall's readability index will be affected.

A. Increase Readability Score of WSDL Document

To provide an example of how to increase the readability score of a WSDL document, we redesign terms in the WSDL document of Amazon's Web service.

Change Domain Terms in WSDL to Non-domain Terms

Domain terms found in WSDL elements may be redesigned to non-domain terms as shown in Table VIII. After redesigning terms, the readability score is recalculated and new relevant scores are obtained in Fig. 4. The new readability score is increased.

Change Domain Terms in WSDL to Those That Are Shallower in Concept Hierarchy

Domain terms found in WSDL elements may be redesigned to those that are shallower in the concept hierarchy as shown in Table IX. After redesigning terms, the readability score is recalculated and new relevant scores are obtained in Fig. 5. The new readability score is increased.

TABLE VIII
CHANGE OF DOMAIN TERMS TO NON-DOMAIN TERMS IN AMAZON'S WSDL

WSDL Element	Original Domain Term	Redesigned Term
AWSECommerceService	Service	Work
SimilarProducts	Product	Goods
Model	Model	Form
NonNegativeIntegerWithUnits	Integer	Number
BrowseNodes	Node	Object

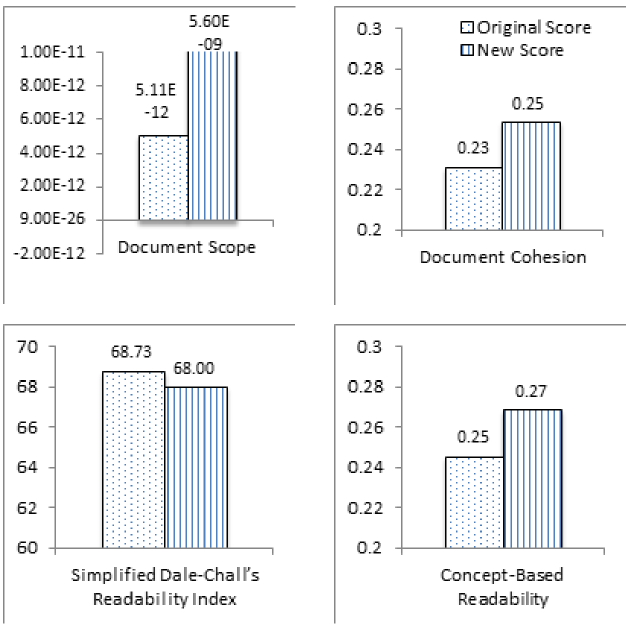


Fig. 4. Adjustment result – Change domain terms to non-domain terms.

TABLE IX
CHANGE TO SHALLOWER DOMAIN TERMS IN AMAZON'S WSDL

WSDL Element	Original Domain Term	Original Depth	Redesigned Domain Term	New Depth
ItemSearch	Item	3	Product	1
Model	Model	2	Type	1
AWSECommerce Service	Service	2	Product	1

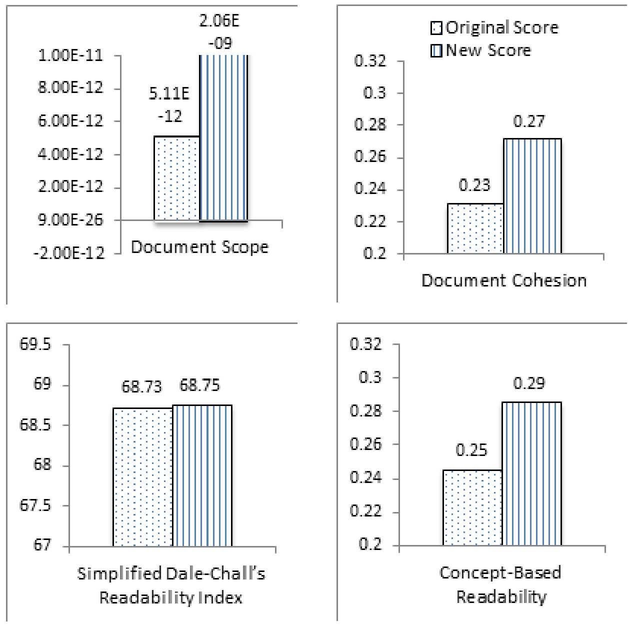


Fig. 5. Adjustment result – Change domain terms to shallower domain terms.

Change or Add Domain Terms So That WSDL Contains Domain Terms That Are More Cohesive with Respect to Concept Hierarchy

Domain terms found in WSDL elements may be redesigned so that they are more closely associated with each other with respect to the concept hierarchy, i.e., the shortest paths between them are shorter. The redesigned terms are shown in Table X and new associations in Table XI. After redesigning terms, the readability score is recalculated and new relevant scores are obtained in Fig. 6. The new readability score is increased.

Change or Add Terms So That WSDL Contains More of Dale's Familiar Words

Domain terms found in WSDL elements may be redesigned to terms that are in the list of Dale's familiar words as shown in Table XII. After redesigning terms, the readability score is recalculated and new relevant scores are obtained in Fig. 7. The new readability score is increased.

Summary

In Fig. 8, we summarize the new readability scores obtained from the adjustment approaches to increase readability. It is likely that change to shallower and more cohesive domain terms has a strong positive impact on readability.

TABLE X
CHANGE TO MORE COHESIVE DOMAIN TERMS IN AMAZON'S WSDL

WSDL Element	Original Domain Term	Redesigned Domain Term
ItemSearch	Item	Product
Model	Model	Type
AWSECommerceService	Service	Product

TABLE XI
SHORTER SHORTEST PATHS BETWEEN PAIRS OF DOMAIN TERMS

Original Pair	Original Length	Redesigned Pair	New Length
Item-Sale	7	Product-Sale	6
Item-Value	6	Product-Value	5
Item-Unit	6	Product-Value	6
Model-Sale	6	Type-Sale	4
Model-Value	5	Type-Value	3

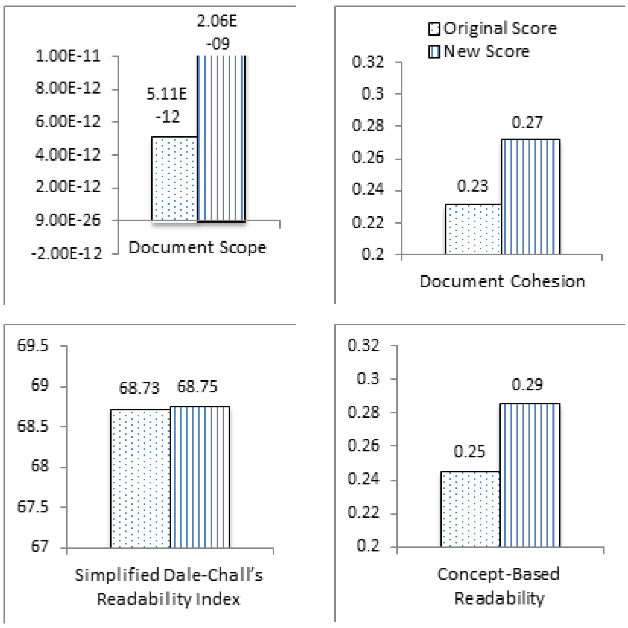


Fig. 6. Adjustment result – Change domain terms to more cohesive domain terms.

TABLE XII
CHANGE TO TERMS AMONG DALE'S FAMILIAR WORDS IN AMAZON'S WSDL

WSDL Element	Original Term	Redesigned Term
ItemDimensions	Item	Whole
BinParameter	Parameter	Writing
Request	Request	Call
Brand	Brand	Name
BrowseNode	Browse	Search

B. *Lower Readability Score of WSDL Document*

To provide an example of how to lower the readability score of a WSDL document, again we redesign terms in the WSDL document of Amazon's Web service.

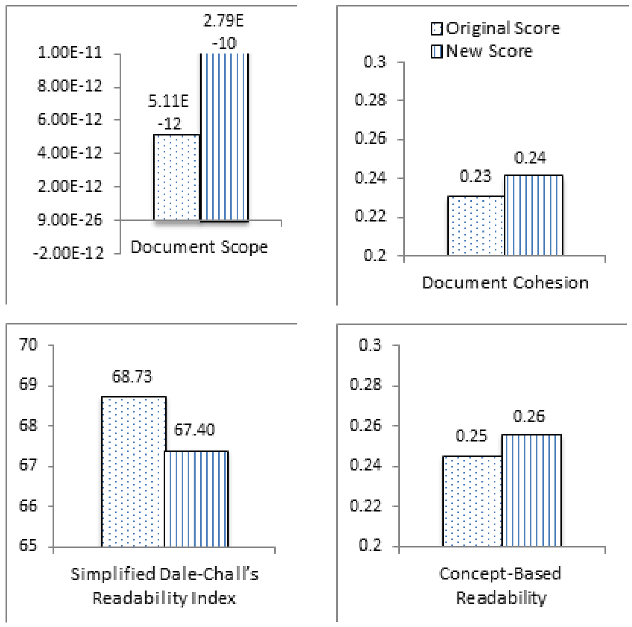


Fig. 7. Adjustment result – Change to terms among Dale's familiar words.

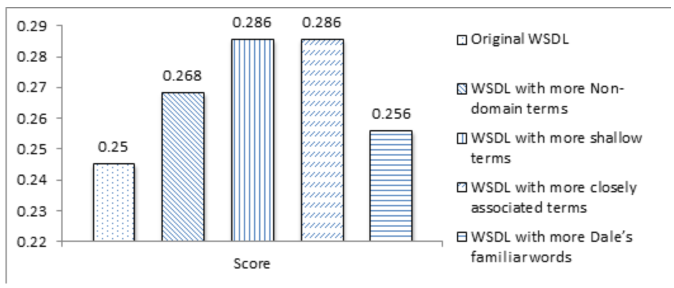


Fig. 8. Comparison of adjustment approaches to increase readability score.

Change Non-domain Terms in WSDL to Domain Terms

Non-domain terms found in WSDL elements may be redesigned to domain terms as shown in Table XIII. After redesigning terms, the readability score is recalculated and new relevant scores are obtained in Fig. 9. Note that, in this experiment, the change of terms can decrease the document scope but at the same time it increases the document cohesion and simplified Dale-Chall's readability index, resulting in higher overall readability.

TABLE XIII
CHANGE OF NON-DOMAIN TERMS TO DOMAIN TERMS IN AMAZON'S WSDL

WSDL Element	Original Non-domain Term	Redesigned Domain Term
Detail	Detail	Specification
Shipping	Shipping	Delivery
Feature	Feature	Function
PackageDimensions	Package	Parcel
IsCategoryRoot	Category	Entity

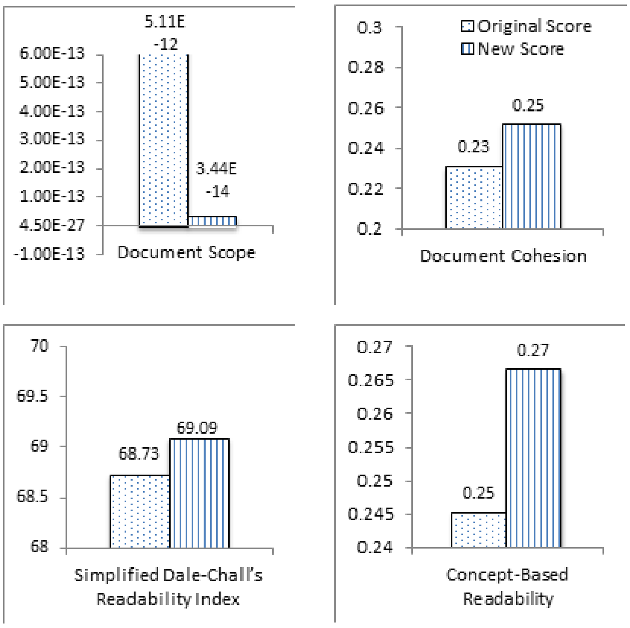


Fig. 9. Adjustment result – Change non-domain terms to domain terms.

Change Domain Terms in WSDL to Those That Are Deeper in Concept Hierarchy

Domain terms found in WSDL elements may be redesigned to those that are deeper in the concept hierarchy as shown in Table XIV. After redesigning terms, the readability score is recalculated and new relevant scores are obtained in Fig. 10. The new readability score is lower.

Change or Add Domain Terms So That WSDL Contains Domain Terms That Are Less Cohesive with Respect to Concept Hierarchy

Domain terms found in WSDL elements may be redesigned so that they are more loosely associated with each other with respect to the concept hierarchy, i.e., the shortest paths between them are longer. The redesigned terms are shown in Table XV and new associations in Table XVI. After redesigning terms, the readability score is recalculated and new relevant scores are obtained in Fig. 11. The new readability score is lower.

Change or Add Terms So That WSDL Contains Less of Dale's Familiar Words

Domain terms found in WSDL elements may be redesigned to terms that are not in the list of Dale's familiar words as shown in Table XVII. After redesigning terms, the readability score is recalculated and new relevant scores are obtained in Fig. 12. The new readability score is lower.

Summary

In Fig. 13, we summarize the new readability scores obtained from the adjustment approaches to lower readability. Compared to the original readability score, the readability scores obtained from three approaches decrease as expected. However, the readability score of the approach in which more domain terms are added increases. This is possible because change of domain terms always affect all three components of the readability model. In this example as seen previously in Fig. 9, the added domain terms that help decrease the document scope also have positive effects on both document cohesion and simplified Dale-Chall's readability index. This hence results in an increase in readability. It is seen that selecting the right terms to remove from or add to the WSDL document may not be easy and a single adjustment approach may not achieve the desired effect.

TABLE XIV
CHANGE TO DEEPER DOMAIN TERMS IN AMAZON'S WSDL

WSDL Element	Original Domain Term	Original Depth	Redesigned Domain Term	New Depth
Brand	Brand	1	Entity	2
BrowseNode	Node	1	Location	2
IssuesPerYear	Per	1	Individual	3
Offers	Offer	1	Provisioning	2

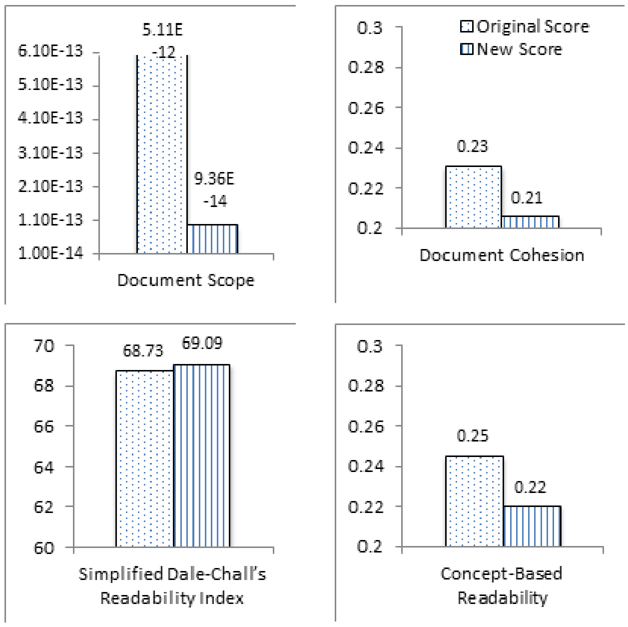


Fig. 10. Adjustment result – Change domain terms to deeper domain terms.

TABLE XV
CHANGE TO LESS COHESIVE DOMAIN TERMS IN AMAZON'S WSDL

WSDL Element	Original Domain Term	Redesigned Domain Term
Brand	Brand	Entity
BrowseNode	Node	Location
IssuesPerYear	Per	Individual
Offers	Offer	Provisioning

TABLE XVI
LONGER SHORTEST PATHS BETWEEN PAIRS OF DOMAIN TERMS

Original Pair	Original Length	Redesigned Pair	New Length
Brand-Offer	3	Entity-Provisioning	4
Node-Value	3	Location-Value	5
Per-HTTP	2	Individual-HTTP	6
Offer-Hour	3	Provisioning-Hour	4
Offer-Warranty	3	Provisioning-Warranty	4

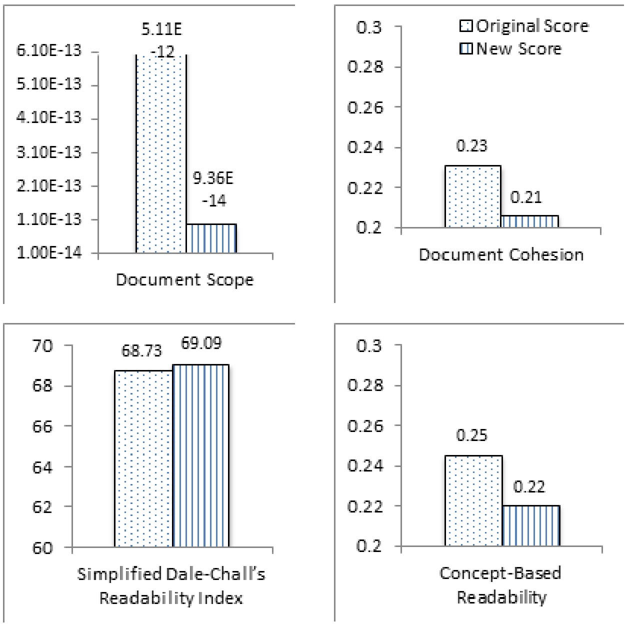


Fig. 11. Adjustment result – Change domain terms to less cohesive domain terms.

TABLE XVII
CHANGE TO TERMS NOT AMONG DALE'S FAMILIAR WORDS IN AMAZON'S WSDL

WSDL Element	Original Term	Redesigned Term
Name	Name	Brand
SearchBinSet	Search	Browse
NarrowBy	Narrow	Specify
CartGet	Get	Acquire
Length	Length	Duration

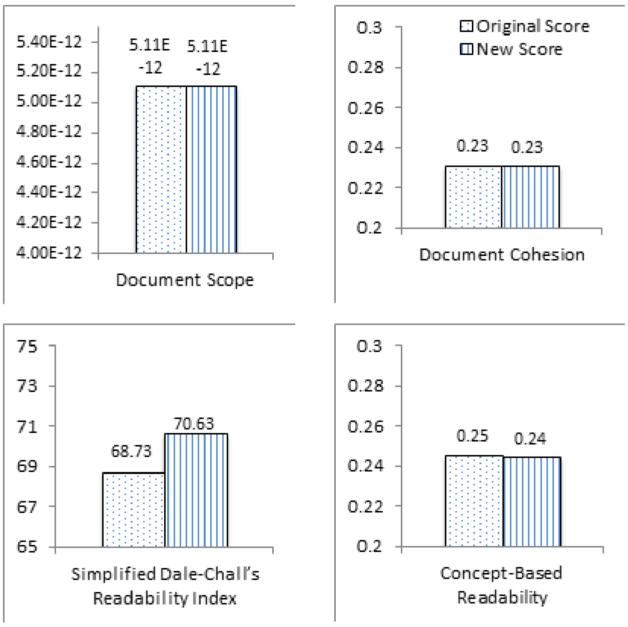


Fig. 12. Adjustment result – Change to terms not among Dale's familiar words.

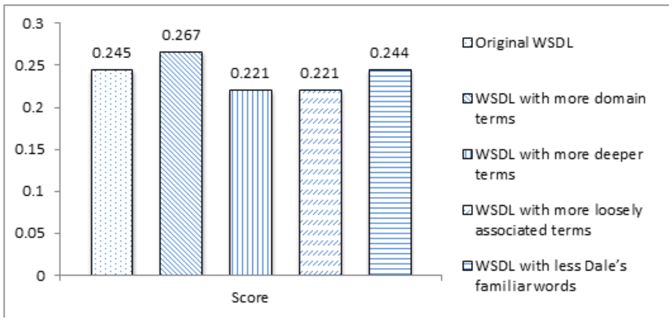


Fig. 13. Comparison of adjustment approaches to lower readability score.

VII. DISCUSSION AND CONCLUSION

This paper presents an application of a concept-based document readability model to measure readability of Web services descriptions. The approach is motivated by a design principle to expose readable service information and by potential risks from service imitation and attacks which may follow such information exposure. The main components of the model are the scope and cohesion of a WSDL document with regard to the vocabulary of its service domain, as well as the difficulty of terms that are used in the document.

Readability can be one factor to service popularity and service imitation and attacks, but it does not say that a highly readable service description will make the service more popular among service consumers, or become a target of competitors and attackers. A WSDL document with high readability score may not attract service consumers if other kinds of service quality are not well-maintained. Likewise, the service provider may also take additional measures to handle vulnerability from exposure of highly readable information.

It is important to note that the readability score of a WSDL document can change if the service domain ontology and the list of familiar words are changed. Several ontologies of different quality may be published on the Internet for a particular service domain, and thus the assessor should pick the domain ontology that is most complete and from a reliable source. We also see that the use of the familiar word list is a useful intuitive approach. Despite the fact that the Dale's list contains familiar words that are known in reading by at least 80 percent of the children in Grade 4, the terms familiar and unfamiliar describing the words here are used in a statistical sense. The percentage of words outside this list is a very good index of the difficulty of reading materials, and has so far been used to determine readability in various contexts including that of technical materials. Therefore we still use the Dale's list in the model even though we see the possibility that an extended list, constructed for readers at higher grade levels, should reflect better the readability scores of technical WSDL documents whose readers are programmers.

Readability adjustment requires that the assessor or service designer give careful consideration to the context of WSDL elements before redesigning them with new terms. It may not be easy to select new terms that are appropriate and give the expected change to the readability level.

Our future work is to experiment with a number of service providing organizations in Thailand to see if the approach can help them with the design of service descriptions and how they use the adjustment method to adjust readability of their WSDL documents. The result from this experiment can be used to determine the appropriate WSDL design which gives the clients a satisfactory readability level. In addition, the readability score and readability adjustment method should be validated. To do so, a number of service clients will be asked to answer a questionnaire to validate if the score truly reflects WSDL readability and the adjustment method can truly increase or lower readability.

REFERENCES

[1] World Wide Web Consortium (W3C). (2001, Mar. 15). *Web Services Description Language (WSDL) 1.1* [Online]. Available: http://www.w3.org/TR/wsdl

[2] J. J. Pikulski. (2013, Aug.). *Readability*. Houghton Mifflin Harcourt. [Online]. Available: http://www.eduplace.com/state/author/pikulski.pdf

[3] X. Yan, D. Song, and X. Li, "Concept-based document readability in domain specific information retrieval," in *Proc. 15th ACM Int. Conf. Information and Knowledge Management (CIKM 2006),* 2006, pp. 540-549.

[4] T. Gruber. (2013, Aug.). *Ontology* [Online]. Available: http://tomgruber.org/writing/ontology-definition-2007.htm

[5] World Wide Web Consortium (W3C). (2004, Feb. 10). *OWL Web Ontology Language Overview* [Online]. Available: http://www.w3.org/TR/owl-features/

[6] J. Zhao and M. Y. Kan, "Domain-specific iterative readability computation," in *Proc. 10th Annual Joint Conf. Digital Libraries (JDCL 2010)*, 2010, pp. 205-214.

[7] A. Jatowt, K. Akamatsu, N. Pattanasri, and K. Tanaka, "Towards more readable web: measuring readability of web pages based on link structure," *SIGWEB Newsletter*, Winter 2012, pp. 1-7.

[8] Y. J. An, J. Geller, Y. T. Wu, and S. A. Chun, "Automatic generation of ontology from the deep web," in *Proc. 18th Int. Workshop Database and Expert Systems Applications (DEXA 2007)*, 2007, pp. 470-474.

[9] Princeton University. (2013, Aug.). *WordNet A Lexical Database for English* [Online]. Available: http://wordnet.princeton.edu/

[10] J. S. Chall and E. Dale, *Readability Revisited: The New Dale-Chall Readability Formula*. Brookline Books/Lumen Editions, 1995.

[11] Stanford Center for Biomedical Informatics Research. (2013, Aug.). *Protégé Ontology Library* [Online]. Available: http://protegewiki.stanford.edu/wiki/Protege_Ontology_Library

[12] DARPA. (2013, Aug.). *DAML Ontology Library* [Online]. Available: http://www.daml.org/ontologies/

[13] Open Knowledge Foundation (OKFN). (2013, Aug.). *Linked Open Vocabularies (LOV)* [Online]. Available: http://lov.okfn.org/dataset/lov/index.html

[14] Stanford Center for Biomedical Informatics Research. (2013, Aug.). *The Protégé Editor and Knowledge Acquisition System* [Online]. Available: http://protege.stanford.edu

[15] Eclipse Web Tools Platform Project. (2013, Aug.). *Eclipse Java EE IDE for Web Developers* [Online]. Available: http://www.eclipse.org/webtools

[16] Amazon Web Services LLC. (2013, Aug.). *eCommerce Web Service* [Online]. Available: http://webservices.amazon.com/AWSECommerceService/AWSECommerceService.wsdl

[17] PayPal. (2013, Aug.). *Invoicing Service API* [Online]. Available: https://svcs.paypal.com/Invoice/CreateAndSendInvoice?wsdl

[18] eBay Inc. (2013, Aug.). *eBay Shopping API* [Online]. Available: http://developer.ebay.com/webservices/latest/ShoppingService.wsdl

[19] E-Business and Web Science Research Group. (2011, Oct. 1). *GoodRelations Language Reference Version 1.0* [Online], Available: http://www.heppnetz.de/ontologies/goodrelations/v1.html

Pananya Sripairojthikoon has been working as a software engineer for Accenture Thailand since June 2012. She received B.Eng. in Computer Engineering from King Mongkut University of Technology Thonburi, Thailand in 2010, and M.Sc. in Software Engineering from Chulalongkorn University, Thailand in 2013. Her research interest includes service computing and application of semantic technology. This research was conducted while she pursued a Master degree.

Twittie Senivongse is an associate professor at the Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Thailand. She received B.Sc. in Statistics from Chulalongkorn University in 1989, M.Sc. in Computing Science from Imperial College, UK in 1992, and Ph.D. in Computer Science from University of Kent, UK in 1997. Her research interest includes service computing and application of semantic technology.

An Empirical Assessment of Quasi-Permanently Vacant Channels in Mobile Communication Bands for Cognitive Radio

Neeta Nathani*, **Gopal Chand Manna****, **S.B.Mule***

**Research Scholar, Department of Electronics Engineering, G.H. Raisoni College of Engineering, Nagpur, India.*

***Senior General Manager, Indian Telecommunication Service, BSNL, Jabalpur, India.*

neeta_nathani@yahoo.com, gcmanna@gmail.com,mulesb1@gmail.com

Abstract- **Basis of cognitive radio is to exploit unused frequency channels in licensed band. Recently standardized IEEE 802.22 set of cognitive radio protocols envisages fixed and nomadic receivers at below 800 MHz bands. Radio link design for this cognitive radio consider that the channels are available only dynamically to secondary users. Scanning period is thus embedded in link layer control as overhead and reduces overall efficiency of cognitive radio technology. For mobile receivers, availability of permanent channels for radio link control is essential for in-band signaling. Existing mobile communication system uses CDMA 800, GSM 900, GSM 1800 and WCDMA 2000 MHz bands for which approximately 1/8th of the band capacity is used for in-band signaling. Present work provides an assessment of vacant channels in mobile communications range which were permanently available at the time of measurement. The study used conventional Radio frequency scanners available for different bands and dedicated engineering handsets for tracking active frequencies. The tests were performed with assemblies carried in a vehicle and across the length and breadth of each city. The allotment of frequencies by Frequency Regulatory Authority to local mobile operators along with the frequency bands reserved for further distribution were also considered. The experimentally collected data were analyzed using RF analysis software and spread sheet database. An analysis of the collected data lead to arrive at the conclusion that more than 1/8th part of resources of each band are nearly permanently vacant which is enough to design in-band common control signalling methods for cognitive radio. Also, with the help of collected data of eight cities, an empirical relationship has been established which can be taken as a thumb rule for projection of channel occupancy from densely populated core areas of big cities to less populated township areas when population is known.**

Keywords: **Cognitive Radio, Mobile communication, In-band signalling, Radio link control, common control signalling.**

Manuscript received October 5, 2013.

Neeta Nathani is with the Department of Electronics Engineering as a Research Scholar in G.H.Raisoni College of Engineering, Nagpur, Maharashtra, India (Tel: 91-9584006543; email: neeta_nathani@yahoo.com).

Dr.G.C. Manna is with the Department of Telecommunications as a Senior General Manager in Inspection Circle, BSNL, Jabalpur, Madhya Pradesh, India (Tel: 91-9425803500; fax: 91-761-2678592; email: gcmanna@gmail.com)

S.B.Mule is with the Department of Electronics Engineering, as a Research Scholar in G.H.Raisoni College of Engineering, Nagpur, Maharashtra, India (Tel: 91-9822427815; email: mulesb1@gmail.com)

I. INTRODUCTION

THE continued success of mobile wireless industry is placing more demands on an already scarce and valuable radio frequency resources banking on commercially successful technologies. Regulators of every country put an eye to oncoming recommendations arising out of standardization institutes and mark corresponding blocks as reserved for future licensing. This has been added up with continued demand for higher bandwidth due to data and video. As a result, one most promising technology has come up for efficient spectrum utilization of the unused channels in the licensed band termed as Cognitive Radio (CR). Cognitive Radio users are termed as secondary users whereas licensed users are termed as primary users. Since secondary user uses the same band as that of a primary user, the former has to take care that the later one is not (or minimum) disturbed in terms of access and interference. CDMA in 800 MHz band, GSM at 900MHz and 1800 MHz bands, WCDMA in 2000MHz band, LTE in 2300 MHz band and WiMAX at 2500MHz and 3500MHz bands are dominant players in mobile communication radio. Wi-Fi and other low power radios play at ISM band at 2400MHz band. To give space to a demanding frequency channel of a band to a primary user, secondary user has to keep track of all the vacant channels of all bands and their respective qualities in terms of SNR and throughput. Thus, cognitive radio users deploying opportunistic spectrum access has to scan continuously the whole range of spectrum with a reasonable bandwidth and scanning frequency. This is a very cumbersome and computing intensive mechanism. Alternatively, a scanner which scans only predefined channels of specified bandwidths and across licensed mobile radio frequency bands only, shall be very efficient and trustworthy. Matched filtering and coherent detection technique is most suitable in such environment for it's short detection time, high SNR, low probability of missed detection and false alarm [1]. Matched filtering closely resemble to SDR compatible Mobile Station (MS) available commercially and also in digital RF scanner used for drive test purpose. Also, most of the communication system needs a pilot and dedicated control channels which remains as fixed during the entire period of communication between two users. This needs association of a permanent frequency channel with cognitive radio, preferably in-band hired channel along-with primary users or select a channel out of band. In opportunistic spectrum access, secondary users will use licensed spectrum when primary users do not transmit. According to Beibei Wang and K J Ray Liu, availability of such a global control channel is not realistic since there may be no permanent channel available for secondary users [2]. Probability to get an opportunistic control channel in licensed band for all nodes in cognitive radio network has been narrated as dramatically small by Antonio De Domenico and co-authors [3]. Thus, exploration for availability of Quasi-permanent channels in licensed mobile communication bands for cognitive radio shall truly facilitate with the advantages of working with common control channel in MAC layer. In the present study, the licenses allocated to mobile service providers of India has been taken into consideration for estimation of grossly available vacant channels of reserve bands and drive tests were conducted in eight different cities for completely allocated GSM bands. It was established that more than one-eighth part of all bands are vacant which closely resembles to the requirement of one signaling channel for seven traffic channels.

The remainder of the paper is organized as follows: Section II explains the signaling channel part requirement as a fraction of channel capacity for mobile communication technology systems which enable us to draw empirical statement for access of the whole licensed band by cognitive radio. Section III covers literature survey indicating the availability of channels in different commercial bands. Section IV discusses about the reserve spectrum and actual allotment of channels by the spectrum allocation authority. Section V describes the measurement plan and collection of data at different cities. This is followed by presentation and discussion of the collected results in Section VI. Finally, the conclusions are presented in Section VII.

II. EMPIRICAL STATEMENT

Regulatory Authorities of every country identify frequency bands for public mobile communication technologies e.g. CDMA, GSM, WCDMA, WiMAX etc. Each band is divided into radio frequency (RF) channels for license purpose. Based on technology deployed, operators divide each RF channel into communication channels which consists of common control (CC) channels and data (traffic) channels. Common Control channels services traffic channels and are within licensed RF i.e. in-band. In 800 MHz band, CDMA IS-95 uses 1.25MHz bandwidth as communication channel with one pilot and seven common control channels i.e. total 8 channels out of total 64 Walsh code channels. Walsh code count was 128 for CDMA 2000 1x i.e. one RF. In 1x EV- DO there are 8 control channel time slots out of total 256 time slots and 96 chips of pilot is embedded in each data slot of 1024 chips which together counts for $1/8^{th}$ part of the channel. In WCDMA, out of total radiated power of 20Watts, primary and secondary control channels and pilot channels constitutes 2.585 Watts leaving 17.415 Watts for traffic. In GSM, if there is one RF, the first time slot works as the pilot and common control channel for remaining 7 time slots. In case of 2 or more RFs, on an average, one time slot of first RF is used as control channel for each additional RF. In OFDMA, for a 5 MHz channel, there are 60 pilot sub-carriers for 360 data sub-carrier and 92 null sub-carriers and thus the ratio is slightly higher than $1/8^{th}$ part of the whole RF channel.

A: Necessary Condition : Designers of mobile network communication distributes the communication access system equipments based on population density at different places. The population density has been identified as dense-urban, urban, sub-urban, highway and rural. Conventional 3 sectored antenna system is used for all cell sites except highway sites where only 2 sectored antenna is deployed. We take into consideration urban areas where occupancy of deployed channels are considerably higher. (i) It shall be fair enough to estimate that a minimum of 30% of the channels are occupied in the sub-urban areas even if operators uses them inefficiently. (ii) Occupancy grows linearly with population at lower population areas and the efficiency of channel utilization also increases. (iii) With increase of population occupancy of channels increases, operators use different optimization techniques for maximum use of available channels but slowly attains saturation. The increased need of channels with population increment and counter-balancing the need through efficient optimization techniques can be mathematically expressed as negative requirement of channels and graphically it can be expressed as saturation. With further increase of population, operators has to borrow traffic channels from higher frequency band lowering the occupancy at lower frequency band and consequently it may be reflected as occupancy depression in lower band. Thus, the occupancy of channels with increase of population can empirically be expressed by the relation: $y = 30+15x-x^2$ where; y= percentage occupancy of channels in lower frequency band and x= population size in millions.

B. Sufficient Condition : Approximately $7/8^{th}$ part of a licensed RF channel of a frequency band is available for useful communication and the remaining part is used for control purpose. As a corollary of the above, we can say that $7/8^{th}$ part of a band is used for purposeful data communication and remaining $1/8^{th}$ part for common control and signaling purpose. Alternatively, if $1/8^{th}$ part of a frequency band is permanently available as spare and the same is used for common control purpose, the total band can be completely accessed for cognitive radio communication purpose.

III. LITERATURE SURVEY

An extensive measurement campaign was conducted in the city of Aachen located in Germany near to the border to Netherlands and Belgium. The spectrum occupancy of several frequency bands was measured over longer time periods up to seven days. The Authors deployed the amplitude probability distribution method to decide spectrum occupancy data which can help to optimize the search for unused spectrum i.e. white space[4]. The Authors identified FM radio bands around 100MHz and the GSM channels bands around 900MHz as the measured minimum level was above -80 dBm. Marc McHenry did large scale spectrum occupancy measurement and observed that the spectrum occupancy in several American cities was found to be always below 25%, which is mostly caused by a rather high decision threshold between -90 and -105 dBm [5]. In Qatar, A comprehensive wideband spectrum occupancy measurement was performed over multiple dimension (time, frequency and space) concurrently over an utilized bandwidth in the 700 − 3000MHz band over a three day measurement period. Across four different locations, the bandwidth utilization varies between 4% and 15%. It was observed that bandwidth utilization is time variant with peak times ranging from afternoon to sunset in almost all of the locations [6]. In Romania, corresponding to the downlink communication direction i.e. for 880 - 960 GSM, E-GSM, Military bands, occupancy level was found by Authors to be which covered the frequency range from 75MHz to 3GHz in an outdoor urban environment. The measured frequency range was divided into six consecutive 500 MHz blocks. Each block was measured during a continuous period of 48 hours and was observed that only 22.57% of the whole frequency range between 75MHz and 3GHz was occupied. 46.80% and for 1710 - 1880 GSM 1800, occupancy was 22.86% [7]. A survey was conducted in Barcelona, Spain [8] to carry out a broadband spectrum measurement campaign A similar type of survey was conducted in Singapore [9] to find the spectrum usage pattern in frequencies from 80 MHz to 5.85 GHz. The 24-hour measurement had been taken at the roof top of Institute for Infocomm Research's building over 12 weekday periods, which indicated that the average occupancy on the frequency range from 80 MHz to 5.85 GHz was only about 4.54%. In a survey at the Loring Commerce Centre [10], the measurements were made during a normal work week for 3 days for the frequency ranges between 100 MHz to 3 GHz. Based on the results of the study, it was observed that the average spectrum usage during the measurement period was 1.7%.

The static measurement shows that the mobile licensed bands and ISM bands are partly occupied and the remaining parts of the spectrum resembles noise. It is apparent that no study has been made to assess actual occupancy of mobile communication bands based on measurements at different locations of the same city in licensed band. Also, no study has been made to find variation of channel occupancy for different type of cities. These studies shall help us to accurately identify the frequency channels with low or no active utilization and successfully deploy them in cognitive radio technology.

IV. RESERVE SPECTRUM AND ALLOCATION

In India, the radio frequencies are being used for different types of services like mobile communication, broadcasting, radio navigation, satellite communication, defense communication, etc. The National Frequency Allocation Plan (NFAP) forms the basis for development and manufacturing of wireless equipment and spectrum utilization in the country. Frequency bands allocated to various types of mobile communication services in India are mentioned in Table I.

TABLE I

LICENSED BANDS OF DIFFERENT WIRELESS TECHNOLOGIES

Technology used	**Band and Channel Width**	**Technology Bandwidth in MHz**	***Allocation**	***Occupancy/ Remarks**
CDMA	890-915MHz; 935-960MHz; Channel B/W 1.25MHz FDD; 20 channels	25+25	5 Operators- 2 RF each	50% vacant
GSM 900MHz	824-849MHz; 869-894MHz; 124 channels of 200 kHz spacing	25+25	4 Operators	Fully Licensed- vacant channels to find out experimentally
GSM 1800 MHz	1710-1785MHz;1805-1880MHz; 374 channels	75+75	4 Operators	60% vacant channels; actual to find
WCDMA	1920-1980MHz;2110-2170MHz	60+60	3Operators-5 MHz each	75% vacant channels
WiMax & 4G	2500–2690MHz	190	3 Operators- 15 MHz each	77% vacant channels
WiMax (fixed)	3400-3600MHz	200	One -20MHz	90% vacant channels

*As per available information.

So far CR technology has been developed keeping into consideration that all channels in licensed bands are dynamically accessed and released. Game theory is applied for capturing a vacant channel inviting competition amongst CR operators. Further, there is a possibility of occupying a vacant band by CR secondary user in the guise of a primary user. In addition, a concept of bidding for vacant channels has also come up for use by secondary users at a comparative cost. All these factors introduces the concept of business out of the unused channels of a licensed user with or without having a permanent infrastructure for CR. It is thus envisaged that mainly public authorities providing public utility services at almost zero cost may be authorized to operate over unoccupied spectrum even though licensed. However the number of such public utility service providers is expected to be very limited. Table I clearly indicates that one out of total eight communication channel in each band is permanently available as vacant for using as pilot and signalling purpose which is necessary for the working of Cognitive Radio communication except GSM 900 MHz band. Conduction of field test is absolutely necessary for assessment of quasi-permanently vacant channels for use of in-band common control signalling purposes.

V. MEASUREMENT SETUP AND DATA COLLECTION

The measurement setup used is a drive test equipment which performs tests in a cellular network and collects data on a moving vehicle. The software and hardware used are: Laptop with charger and USB hub, GPS and data cable, Digital RF Scanner, License dongle for TEMS, Cell site database and Link budget, Clutter diagram from Google website, MapInfo software, Engineering handsets with 4 (2G/3G) SIMs of different operators mounted simultaneously and cable terminal. In the setup, data collection software is installed in the laptop which uses a mobile phone along with GPS system. Drive test route was carefully selected and data was collected accordingly. GPS collects the data of latitude and longitude of each point while mobile collects data like signal strength, uplink and downlink frequency, etc. All the information is stored with their respective date and time as well as their geographical locations. File sizes are kept within limit of about 25MB in each part trace and a new file is open in each part. Data from all part files are grouped from collection software and are stored in some output file. Using the above setup, measurement was taken on eight cities at different geographical locations viz. Kolkata, Patna, Ranchi, Dibrugarh, Jaipur, Shillong, Bhopal and Port Blair which have population ranging from 1.5 million to 6.6 million shown in Table 2.

Figure 1 shows the measurement route map of Jaipur. Jaipur city has a population of about 6.66 million. One side of measurement route is heavily populated while it is less populated on the other side. The lower band of frequency channels is distributed among three operators while the upper band is allocated to one operator. While moving from location having 75.82894^0E and 26.91442^0N, it is found that in lower band of Operator1 -14 BCCH are occupied namely 71,73,75,76,77,78,80,81,82,87,88,111,112,113 while 29 BCCH are vacant (72,74,79,83 to 86,89 to 110) of which BCCH no. 89 to 110 is not used at all. Among hopping channels, two hopping channels 114 and 121 are vacant. While in Operator2, 13 BCCH (1 to 13) and 18 hopping channels (14 to 31) are fully occupied. In Operator3, 13 BCCH (50 to 62) and 18 hopping channels (32 to 49) are also fully occupied .The frequency band is switched to upper band used by Operator 4, in which 15 BCCH (530 to 544) and 15 hopping channels (546 to 560) are occupied while one BCCH number 545 is vacant. So, on an average 75% of frequency spectrum of lower band is utilized.

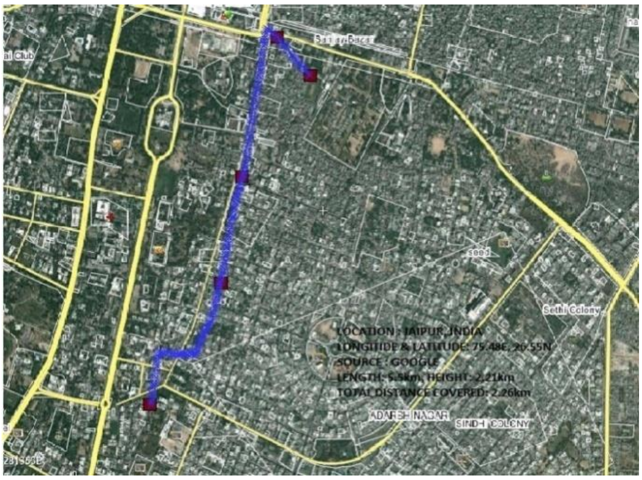


Fig. 1. Measurement Route map of Jaipur

Kolkata is a densely populated city which has a population of about 4.48 million in core Municipal area. During the survey, the area selected was busiest one which has tall buildings towards right side at North and scattered single storied structure towards the river. Towards South, it is almost green field; measurement route is shown in Figure 2, it is found that the lower band and upper band of frequency channels is distributed among three operators. In the two operators of lower band in Kolkata all the BCCH viz. 70, 71,73 to 76, 78,80,82,87,88,111,112,113 in Operator 1 and 50 to 62 in Operator 2 are fully occupied but 30 BCCH i.e. 72,77,79,81,83 to 86, 89 to 110 are vacant in third operator; and 4 hopping channels i.e., 40,48,49,96 are vacant. In the upper band only one BCCH, i.e., 873 is vacant among the three operators; while no hopping channels are available. In Kolkata, it is found that average spectrum occupancy is about 72.58%.

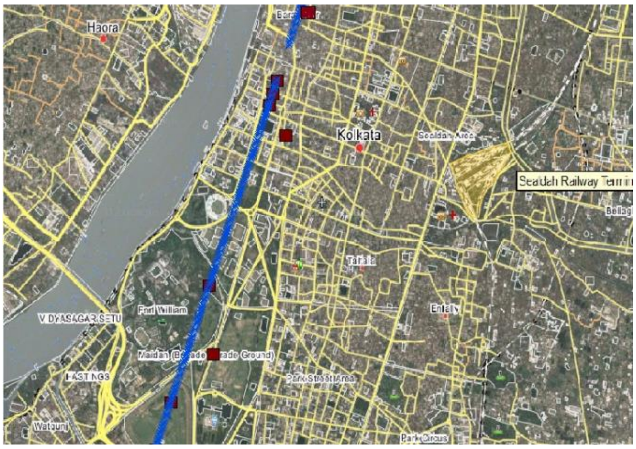


Fig. 2. Measurement Route map of Kolkata

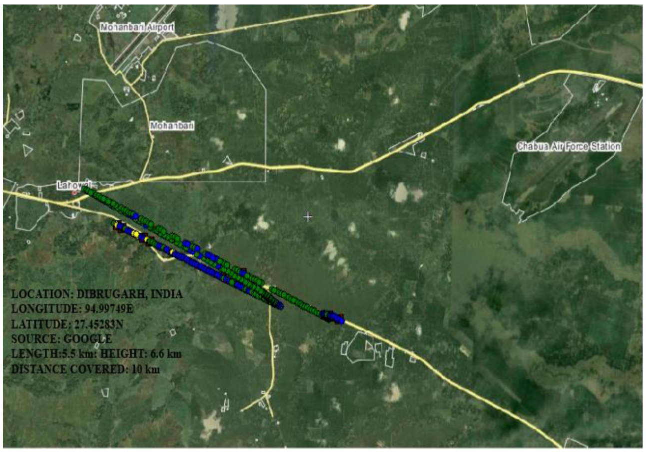


Fig. 3. Measurement Route map of Dibrugarh, Assam

Figure 3 shows the measurement route map of Dibrugarh. Dibrugarh district, well known as the "Tea City of India", has a population of about 1.5 million. It is one of the least populated district where the Brahmaputra river flows in its north. The lower band of frequency channels is distributed among three operators while the upper band is allocated to two operators. While moving from location having 94.99749^{0}E and 27.45283^{0}N, it is found that in lower band of Operator1 - 7 BCCH are occupied namely 1,3,4,5,6,7,8 while 3 BCCH are vacant (2,9,10) and none of the hopping channels are vacant. While in Operator2, 4 BCCH (23 to 26) and 5 hopping channels (27 to 31) are fully occupied. In Operator3, 8 BCCH , i,e; 74,76,78,81,87,88,111,113 and 16 hopping channels namely 63,64,66 to 69, 115 to 120, 122,123,124 are fully occupied .The frequency band is switched to upper band used by Operator4, in which 13 BCCH (536,537,538,867 to 870, 872 to 875,747) and 10 hopping channels (876 to 885) are occupied while 18 BCCH are vacant. In the upper band of Operator2 all 9 BCCH and 3 hopping channels (845 to 847) are occupied. So, on an average 50% of frequency spectrum of lower band is utilized.

Data was also collected for spectrum occupancy measurements in GSM at 900MHz bands in an outdoor environment other cities viz. Bhopal, Patna, Ranchi and Dibrugarh, Shillong & Port Blair. In Bhopal, the spectrum occupancy in lower band is 74.19% while in Ranchi it is only 52.42% as it switches to upper band where it has a spectrum occupancy of 83%. The measurements shows that the spectrum occupancy in lower band of Patna is 75.8%. Shillong, capital of the state of Meghalaya of India is located at 25.57° N and 91.88° E on a plateau in the eastern part of the state. The total population of the city is 1.43 million, has a spectrum occupancy of 54%. Port Blair is the capital city of Andaman and Nicobar Islands in India. Located at 11° 40' N and 92° 46' E, Port Blair is the municipal council in the southern part of Andaman, a part of India's Union Territory. Being the least populated place, its spectrum occupancy is 43%.

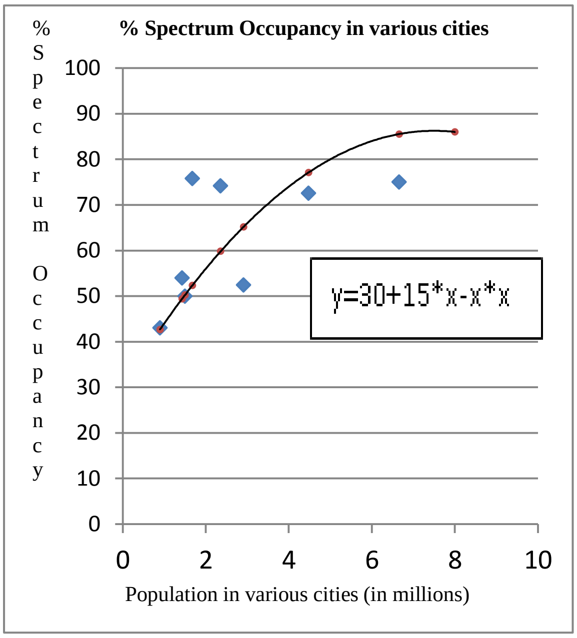


Fig. 4. Population Spectrum Graph

VI. RESULTS AND DISCUSSION

Collected data of eight cities has been recorded in table II along with city population figures. The population data has been collected from internet and limited to city municipal areas only as declared by local civic authority. The population was mostly taken on the basis of 2011 sensus. Figure 4 depicts the spectrum occupancy in eight cities of India In figure 4 above, it is projected that even the most sub-urban area with scanty population shall have nearly 30% occupancy. For population between 1 million to 4 million, the increase is nearly linear. The occupancy saturation is observed in the range of 4 million to 7 million. Also, with projected expansion of densely populated city core areas to 8 million, occupancy level is estimated go upto 86%, leaving a clear space of 14% of channels for use of cognitive radio.

TABLE II POPULATION SPECTRUM CHART

S.No.	City	Population (in millions)	% Spectrum Occupancy
1	Bhopal	2.36	74.19
2	Patna	1.68	75.8
3	Kolkata	4.48	72.58
4	Ranchi	2.91	52.42
5	Jaipur	6.66	75
6	Dibrugarh	1.5	50
7	Shillong	1.43	54
8	Port Blair	0.9	43

The obtained results demonstrated the existence of significant amounts of spectrum potentially available for the future deployment in CR networks in lower bands of GSM which are not at all used.

VII. CONCLUSION

Discussion from above reveals that more than 20% of the completely licensed bandwidth is practically vacant in a saturated market environment. This is well above the requirement of $1/8^{th}$ part of the band to get access to the whole of the bandwidth at a time by cognitive radio and adequate to take additional MAC level overhead required for Cognitive Radio. There is no immediate necessity of these channels by the licensed operators which can be safely deployed as common control channel for cognitive radio purpose. Further, to doubly enhance protection of common control channel, a disaster recovery common control channel may be designated which will hold the replica of allotments and processing status of CR Primary Common Control Channel. Also, depending upon the population, it is possible to approximate channel occupancy and hence cognitive radio technology planners can draw a long term plan for efficient use of them in public benefit. Use of commercially stable technologies with minimum changes at Operator Transceiver equipment and user mobile equipment for adaptation to Cognitive Radio technology shall lead to development of a economically deployable system.

VIII. REFERENCES

[1] Ying Chang Liang, Kwang Cheng Chen, Geoffrey Le Li and Petri Mahonen, "Coginitive *Radio Networking and Communications: An Overview*", in Proc. IEEE Transactions in on Vehicular Technology, 2011.

[2] Beibei Wang and K J Ray Liu, "*Advances in Cognitive Radio Networks: A Survey*", IEEE, February 2011, 1932-4553.

[3] Antonio De Domenico, Emilo Calvanese Strinati and Marie Gabriella Di Benedetto, "*A survey on MAC strategies for Cognitive Radio Networks*", IEEE 2012, 1553-877X/12.

[4] Matthias Wellens, Jin Wu and Petri Mahonen, Department of Wireless Networks, RWTH Aachen University Kackertstrasse, Germany, "*Evaluation of Spectrum Occupancy in Indoor and Outdoor Scenario in the Context of Cognitive Radio*".

[5] M. A. McHenry, P. A. Tenhula, D. McCloskey, D. A. Roberson, and C. S. Hood, "*Chicago Spectrum Occupancy Measurements & Analysis and a Long-term Studies Proposal*," in Proc. of Workshop on Technology and Policy for Accessing Spectrum (TAPAS), Boston, USA, August 2006.

[6] Khalid A. Qaraqe, Hasari Celebi, Ali Gorcin†, Amer El-Saigh‡, Huseyin Arslan†, Mohamed-slim Alouini , "*Empirical Results for Wideband Multidimensional Spectrum Usage*", supported by Qatar Telecom (Qtel)

[7] Alexandru Marțian, Calin Vladeanu, Ioana Marcu, Ion Marghescu, "*Evaluation of Spectrum Occupancy in an Urban Environment in a Cognitive Radio Context*", International Journal on Advances in Telecommunications, vol 3 no 3 & 4, year 2010, http://www.iariajournals.org/telecommunications.

[8] Miguel Lopez-Benitez, Anna Umbert and Fernando Casadevall, "*Evaluation of Spectrum Occupancy in Spain for Cognitive Radio Applications*", IEEE, 2009, vol.no. 978-1-4244-2517-4/09, IEEE Xplore.

[9] Md Habibul Islam, Choo Leng Koh, Ser Wah Oh, Xianming Qing, Yoke Yong Lai, Cavin Wang, Ting-Chang Liang, Bee Eng Toh, Francois Chin, Geok Leng Tan and William Toh, "*Spectrum Survey in Singapore: Occupancy Measurements and Analysis*", IEEE Xplore.

[10] Tugba Erpek, Mark Lofquist and Ken Patton, "*Spectrum Occupancy Measurements Loring Commerce Center Limestone Maine*", September 2007, www. sharedspectrum.com.

Neeta Nathani was born in India in 1981. She received the B.E. and M.Tech. degrees from Rajiv Gandhi Proudyogiki Vishwavidyalaya, India, in 2003,and 2009, respectively. She joined Sri Ram Institute of Technology, Jabalpur, India, in 2003 and worked there for few years. Since 2008, she has been with Gyan Ganga Institute of Technology & Sciences, Jabalpur, India where she is currently an Assistant Professor. Her main areas of research interest are Wireless Communication, Quality of service of Cognitive Radio.

Dr. Gopal Chandra is working as Senior General Manager (Head Quarters), Inspection Circle, BSNL, a wholly owned Company under Department of Telecommunications (DoT), Govt. of India. He has developed and conducted one week course on Quality of Service Monitoring at Information and Communication Technologies Authority, Mauritius as International Expert through Commonwealth Telecom Organisation London during August 2010. He has also deliverd a speech on WiMAX coverage Evaluation at International Conference on Advanced Communications Technology 2011 at Seoul, Korea and chaired a session on Network Management. He had also delivered speech on ADSL at International Telecommunication Union seminar in 2000 at Bangalore, India. He has carried out extensive research on coverage issues of GSM,CDMA, WCDMA and WiMAX radio access. Study of Wireless Traffic and QoS estimation of Cognitive Radio are his current areas of research. In addition, he has written several articles on advanced telecommunications which has been published in national and international journals and symposiums.

Prof. Mule has completed his master degree in Electronics (Computer) from Government College of Engineering, Pune and has six years of industrial experience..

Filtering Performance of Reducing the Sampling Rate of Sound Card: Perspectives on Different Signal-to-Noise Ratios

Bing-Yuh Lu*,**, Li-Man Tseng***, Ling-Yuan Hsu**, Shr-Hong Tang**, Yu-Lin Weng**, Man-Na Hsueh***, Mei-Ju Su****, Huey-Dong Wu*****, Jin-Shin Lai******

*Department of Electronic Engineering, Tungnan University, Taiwan, ROC
**Department of Information Management, Catholic St. Mary's Junior College of Medicine, Nursing, and Management, Taiwan, ROC
***Department of Nursing, Catholic St. Mary's Junior College of Medicine, Nursing, and Management, Taiwan, ROC
****Department of Biomedical Engineering, Yuanpei University, Taiwan, ROC
*****Section of Respiration Therapy, Department of Integrated Diagnostics and Therapeutics, National Taiwan University Hospital, Taiwan, ROC
******Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital and College of Medicine, National Taiwan University, Taiwan, ROC

Benjaminee777@hotmail.com, liman@smc.edu.tw, lyhsu@smc.edu.tw, shrhongtang@smc.edu.tw, ylweng@smc.edu.tw, helena@smc.edu.tw, merri1024@gmail.com, hdwuntuh@ntu.edu.tw, jslai@ntu.edu.tw

***Abstract*— This study presented a theoretical analysis to prove that there exists an optimal sampling frequency to reduce the noise in the signal to reduce the processes of acoustic digital signal processing. The respiration sounds have been recorded by 2 to 44.1 KHz or higher sampling rates, however, the optimal frequency might be decided by the noise from environment. This study solves the problem from the point of theoretical view. The findings were interesting because the noise in a pure sine wave can be reduced by lower sampling rates. The sampling frequency of lung sound recording can be reduced in the range of 2 to 5 KHz. In addition, the results supported to enhance signal-to-noise ratio in the receivers of frequency modulation, under-water acoustics, and other communication applications.**

Manuscript received Aug 20, 2013.
This work was supported in part by the National Science Council, Taiwan under Grant 101-2221-E562-004, 101-2627-E-002-006, and 102-2627-E-002-005.

B. Y. Lu is with the Tungnan University, Taipei, and St. Mary's Junior College of Medicine, Nursing and Management, Yilan, Taiwan. (e-mail: benjamin@smc.edu.tw)
L. M. Tseng is with St. Mary's Junior College of Medicine, Nursing and Management, Yilan, Taiwan. (e-mail: liman@smc.edu.tw)
L.Y. Hsu is with St. Mary's Junior College of Medicine, Nursing and Management, Yilan, Taiwan. (e-mail: helena@smc.edu.tw)
S.H. Tang with St. Mary's Junior College of Medicine, Nursing and Management, Yilan, Taiwan. (phone: +886-922-998305; fax: +886 2 39897398; e-mail: shrhongtang@smc.edu.tw)
Y. L. Weng is with St. Mary's Junior College of Medicine, Nursing and Management, Yilan, Taiwan. (e-mail: ylweng@smc.edu.tw)
M. N. Hsueh is with St. Mary's Junior College of Medicine, Nursing and Management, Yilan, Taiwan. (e-mail: helena@smc.edu.tw)
M. J. Su is with Yuanpei University,HsinChu, Taiwan (Email: merri1024@gmail.com)
H. D. Wu with Section of Respiration Therapy, Department of Integrated Diagnostics and Therapeutics, National Taiwan University Hospital, Taiwan (e-mail: hdwuntuh@ntu.edu.tw)
J. S. Lai with Health Science and Wellness Center, National Taiwan University, Taiwan (e-mail: jslai@ntu.edu.tw)

***Keyword*— noise reduction, sampling frequency, acoustic signal, lung sound, sound card, sonar,SNR.**

I. Introduction

THE linear interpolation technology for recovery of the signal is a common method, but the precision of the interpolation is not clear. In lung sound recording, the suggestive sampling rates are 44.1, 22.05, 11.025, 5.5125, and 2.75625 KHz which are based on the sampling rate of traditional recorders. The sampling is the basic process of the signal processing. The modern lung sound and heart sound recorders, biomedical sensors, digital image systems, storage devices, digital communication systems get the acquisition of raw data from the humans by sampling [1,2]. The computing speeds of the micro-processors are enhanced by the semiconductor manufacturing technology which allowed the were 10 MHz 30 years ago, but that of the advanced communication signal converters are several GHz now [16].

The Computerized Respiratory Sound Analysis (CORSA) [3] pointed out the survey of respiratory sounds research, and very high speed internal clock to control the units of the processors. Therefore, the sampling rate of the electronic devices has been accelerated in the past decades. For example, the limitation of sampling rate of analog-to-digital converters the sampling rate was listed as a parameter of the method of digitalization [4]. The word lengths, and sampling rates were considered to record the lung sound perfectly by using the sound blaster cards or commercial multi-channel signal acquisition cards. The simplest process of the acquired raw data is moving average as a low pass filter to reduce the noise. The algorithm of adaptive filters acquired the environment noise to cancel the noise in the signals [5]. Furthermore, in the computation of the spectrogram, the overlapped data decided its resolution [6]. Therefore, the total error in a specific duration of lung sound can be a dominant factor of precision.

The adjustable sampling rate has been employed to reduce the volume in data storage and communication. Some literatures presented the adjustable sampling rate can enhance the signal-to-noise ratio [7,8]. In some cases, the weak signal with lower sampling rate, and the strong signal with higher sampling rate are the rules to reduce the data volume, and enhance the SNR. Therefore, the experimental results can be discussed by the theoretical analysis. The following proof proposed a direct solution to the optimal sampling rate for reducing the noise.

II. METHOD

If the signal x(t) with noise n(t), and the sampling rate is $\omega_s = \frac{2\pi}{T}$, where T is the period. The sampling pulse train is denoted as *p(t)*. Therefore, the sampled signal is:

$$y_p(t) = \left(x(t) + n(t)\right) g p(t) \quad (1)$$

where $p(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT)$

In this study, the proof gives a solution to reduce noise by taking an optimal sampling frequency.

Taking the Fourier transform [9], we have:

$$Y_p(j\omega) = \left(X(j\omega) + N(j\omega)\right) g P(j\omega) \quad (2)$$

Generally, a low-pass filter with the transfer function of $H(j\omega)$ is always designed as be a pre-processing device to cancel the noise. Therefore, the noise can be optimal reduced by some specific sampling rate in the condition of

$$\frac{d}{d\omega_s}\left(Y(j\omega) - X(j\omega)P(j\omega)H(j\omega)\right) = 0 \quad (3)$$

That is:

$$\frac{d}{d\omega_s}\left(\frac{1}{T}\sum_{k=-\infty}^{\infty} N\left(j(\omega - k\omega_s)\right)\right) = 0 \quad (4)$$

Then

$$\frac{-jk}{T}\sum_{k=-\infty}^{\infty} N'\left(j(\omega - k\omega_s)\right) = 0 \quad (5)$$

The optimal ω_s depends on the solution of the differential of $\sum_{k=-\infty}^{\infty} N'\left(j(\omega - k\omega_s)\right) = 0$. It exists the solution of ω_s to reduce the noise.

To verify the proof, we design a computer simulation to show the statistical view of the reduction of noise. Ideally, the respiration can be simplified as a sine wave whose half cycle for inspiration, and the other for expiration. The noise was the additive white Gaussian noise (AWGN) which is the common noise in the communication system [14]. The parameters of computer simulation were listed in Table 1.

The average absolute errors of the sampling points were computed to evaluate the performance of the method of reducing the sampling rate. The sampling rates were formulized as:

$$f_{si} = \frac{44.1}{2^{i-1}} KHz \quad (6)$$

where i=1 to 5, are also the group indexes, i.e. the sampling frequency of group 1 was 44.1 KHz, and that of group 2 was 22.05 KHz.

The pure signal in group 1 was denoted as $x_1[n]$, and the noised signal was $y_1[n]$ whose numbers of points are N. We reduce the sampling rate by dividing 2^{i-1} where i>1. The numbers of samples were also reduced to be $N/2^{i-1}$. However, the number of the samples can be recovered by well-known linear interpolation technology. Therefore,

$$me_i = \frac{\sum_{k=1}^{k=N} \left|\hat{y}_i - x_1\right|}{N} \quad (7)$$

where $\hat{y}_i$ was the reduced noised signal and recovered by linear interpolation. Consequently, the percentage of error reduction (ξ) can be described as:

$$\xi(\%) = \frac{me_i}{me_1} \cdot 100\% \quad (8)$$

III. RESULTS

The ideally pure sine wave and that with the AWGN and random noise (RN) of signal to noise ratio (SNR) =40 dB were shown in Fig. 1 (a) and (b). The sampling rate was 44.1 KHz. The root-mean-square (RMS) error was computed by different sampling rates of 44.1(Group 1), 22.05(Group 2), 11.025(Group 3), 5.5125(Group 4), and 2.76125(Group 5) KHz, respectively. We run 100 times of computer simulations by the same parameters in Table 1 to compute the mean absolute errors (MAE). The box plots were presented in Fig. 2.

TABLE I THE PARAMETERS of the SIMULATED RESPIRATION SIGNAL.

Variables	Value
Shape	Sine wave
Period	3 sec.
Signal to noise ratio (SNR) of AWGN	20, 30, 40, and 50 dB
Sampling rate	44.1(Group 1), 22.05(Group 2), 11.025(Group 3), 5.5125(Group 4), and 2.76125(Group5) KHz

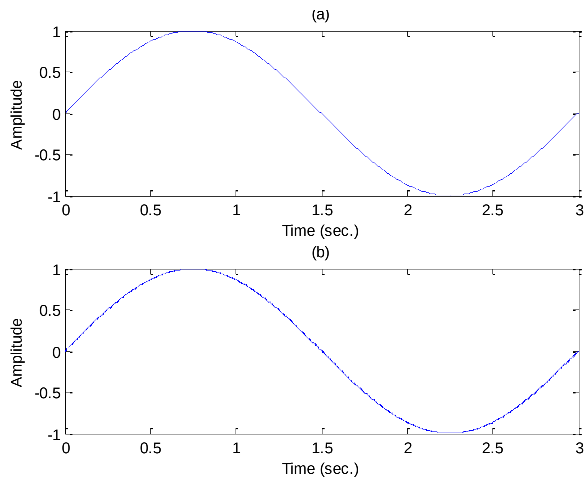


Fig. 1. Computer simulation: (a) the ideally pure sine wave and (b) that with AWGN.

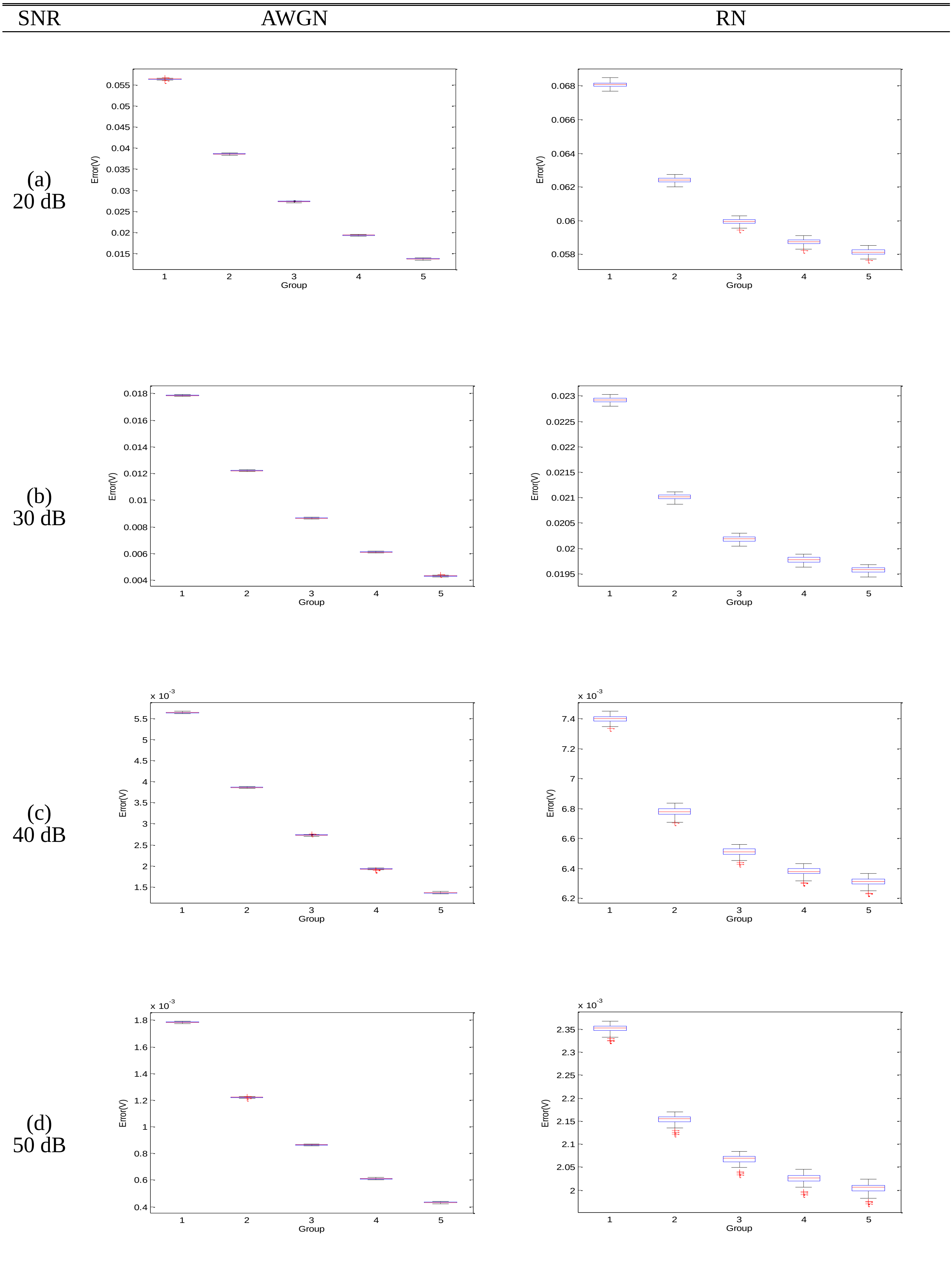


Fig. 2. The box plots of the MAE of Groups 1, 2, 3, 4, and 5 of 100 times. The AWGN level were (a) SNR=20, (b) SNR=30, (c) SNR=40, and (d) SNR=50. (The sampling rates: 44.1(Group 1), 22.05(Group 2), 11.025(Group 3), 5.5125(Group 4), and 2.76125(Group 5) KHz)

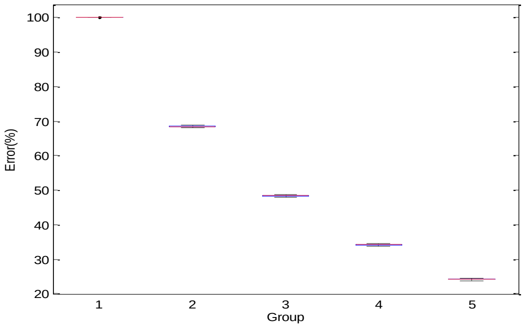
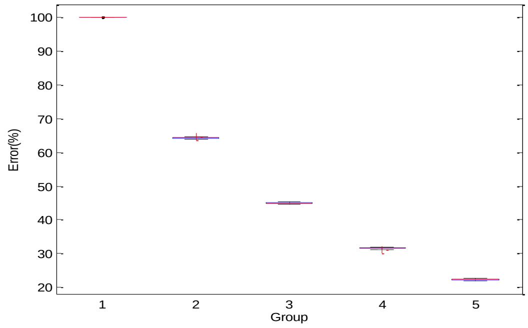
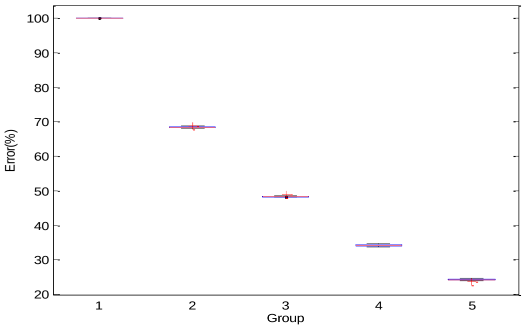
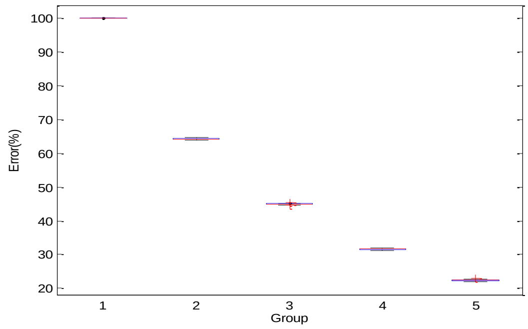
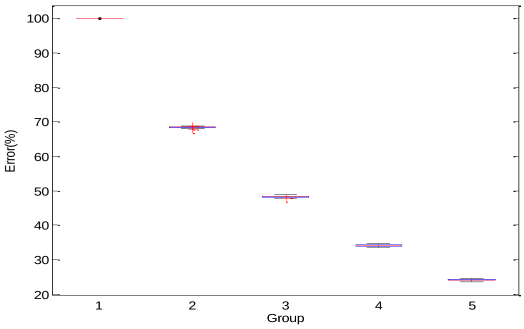
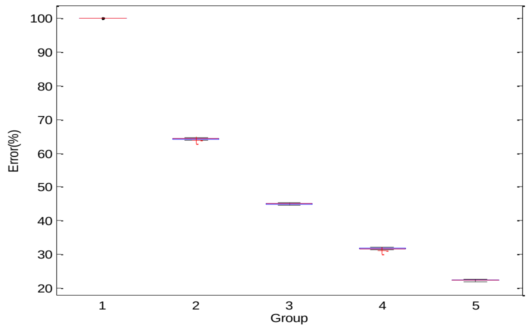
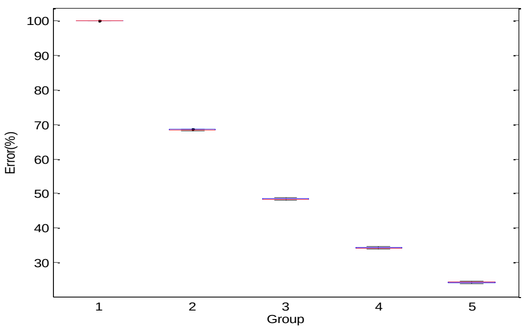
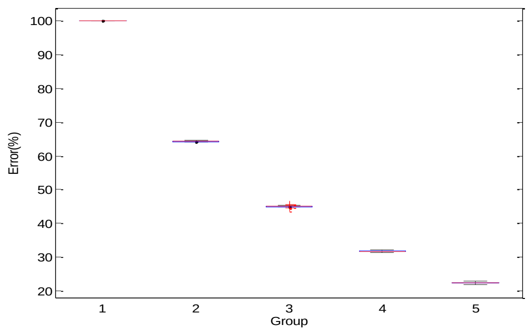
SNR	AWGN	RN
(a) 20 dB		
(b) 30 dB		
(c) 40 dB		
(d) 50 dB		

Fig. 3. The box plots of the percentage of error reduction (ξ) of Groups 1, 2, 3, 4, and 5 of 100 times. The AWGN level were (a) SNR=20, (b) SNR=30, (c) SNR=40, and (d) SNR=50. (The sampling rates: 44.1(Group 1), 22.05(Group 2), 11.025(Group 3), 5.5125(Group 4), and 2.76125(Group 5) KHz)B

Figure 2 presented the statistical results. Each box contained 100 trials of MAE. In (a), the SNR of AWGN is 20 dB. The box plots displayed that Group 5 (sampling rate =2.76125 KHz) has gotten the minimum MAE errors. Compared with the original signal (Group 1), the operation of reducing sampling rate cancelled about 1/5 MAE errors from the point of statistical view. In (b), (c), and (d), the SNRs were 30, 40, and 50 dB. All results show that the smallest sampling rate (Group 5) gets the smallest RMS errors. Based on the proof, there exists a optimal sampling rate to reduce the noise. From the box plots, we find that the lower sampling rate got the lower MAE errors by statistical analysis.

The percentage of error reduction by decreasing the sampling rate was obviously. Figure 3 presented the performance of the reducing sampling frequency. All plots displayed that group 5 achieved the best effect of noise cancellation for both AWGN and RN. The ξ s were in the range of 15 to 25%. Therefore, the filtering performance of the easy method for pure sine waves was excellent.

From the observation of Figs. 2, and 3, the lower sampling rate did reduce both AWGN and RN. The lower frequency signal can be filtered the noises by the lower sampling rate of sound card.

IV. Discussion

The pure sine waves usually play as the carrier in the communication systems [17]. In the detection systems, the reflections of the transmitting pure sine waves are widely employed in sonar, RADAR, and other systems. The sampling rate of receiver can be reduced based on the results of our study, especially in the real-time computing system. Fourier series described that a periodic signal is the linear combination of the fundamental signal and its harmonics. Therefore, the simulated ideally pure sine wave is significant, because of the theory of the linear time invariant (LTI) system [1]. The box plots can be expanded to explain the geniality of the operation of the reducing sampling rate. Chang and Lai [11] have studied on the performance evaluation and enhancement of lung sound recognition system in two real noisy environments. The results presented that the frequencies of noises were higher than that the main episodes in the spectrogram. Therefore, if the sampling rate of the experiment reduced to a specific frequency and processed the raw data with moving average, most of the environmental noise can be much cancelled. However, the experienced medical staff usually records the lung in silent environments whose noises compose with the sounds from air conditioners, fans, computer systems and so on. Based on the sampling theory, the energy of noise would be reduced by the lower sampling rate that shows the possibility of the optimal sampling rate can be found in some specific environments. Furthermore, the processing of the digital data can be accelerated by decreasing the data. The benefits of selecting a better sampling rate can achieve the better performance of signal processing. CORSA indicated that the wheeze was contained in the domain frequency at 400 Hz, but a number of investigators have suggested that the range is actually between 80–1,600 Hz and 350–950 Hz by filter theory [12,13]. Lu et al [14] have synthesized the normal breath and wheezing sounds whose main components were under 2 KHz. All the papers provided that the evidence of the sampling rate can be reduced to cancel the noise. However, the crackle includes the higher frequency components, the optimal sampling rate can be higher than normal breath and wheezing sounds. This is the limitation of using the lower sampling rate for lung sound recording. Furthermore, the standardized sampling rates applied in many industry standard sound facilities are 44.1, 22.05, 11.025 or 5.5125 KHz as standards. Most modern research groups selected 44.1 or 22.05 KHz to record the lung sound. The filtering performances of AWGN and RN were close. Therefore, the filtering method can be employed in the receivers of digital systems to cancel the noises caused from different reasons.

Some electronic stethoscope use the lower frequency of sampling rate [18]. It supports the examination of this study. We give the better choice of sampling rates to the researchers of respiration sounds.

V. Conclusion

The theoretical analysis of reducing sampling rate filter was solved in this study. The proof can be employed to simplify the filter design, if we get the smallest error from the simulated signal by the optimal sampling frequency. In addition the reasons of noises in physiologic signals are usually band-limited, therefore, the optimal sampling frequency can be used to simplify the system design. In the detection systems, the reflections of the transmitting pure sine waves are widely employed in sonar, RADAR, and other systems. The sampling rate of receiver can be reduced based on the results of our study, especially in the real-time computing system.

Acknowledgements

The authors thank for the valuable comments from the reviewers and the supports from the project numbers of NSC101-2221-E-562-004, NSC101-2627-E-002-006, and NSC102-2627-E-002-005, National Science Council, Taiwan, Republic of China.

References

[1] S. Haykin and B. Van Veen, *Signals and systems*, New York: Wiley, 1998.

[2] J. S. Beasley and G.M. Miller, *Modern electronic communication*. New York: Pearson Education, 2005.

[3] A. R. A. Sovijarvi, J. Vanderschoot and J. R. Eavis, *Standard of computerized respiratory sound analysis*. Eur Respir Rev 2000;10:585.

[4] J. E. Earis and B. M. G. Cheetham, "Current methods used for computerized respiratory," *Eur Respir Rev*, vol. 10, 2000 pp. 586-590.

[5] B. S. Lin, H. D. Wu, F. C. Chong and S. J. Chen, "Wheeze recognition based on 2D bilateral filtering of spectrogram," *Biomedical Engineering-Applications, Basis & Communications*, vol 18, 2006, pp. 128-137.

[6] A. R. A. Sovijärvi, L. P. Malmberg, G. Charbonneau, J. Vanderschoot, F. Dalmasso, C. Sacco, M. Rossi and J. E. Earis. "Characteristics of breath sounds and adventitious respiratory sounds," *Eur Respir Rev*, vol. 10, 2000, pp. 591–596.

[7] J. P. McKay and Y. Rahmat-Samii, "Compact range reflector analysis using the plane wave spectrum approach with an adjustable sampling rate," *IEEE Trans. on Ant and Propaga*, vol. 39, 1991, pp.746-753.

[8] A. Hart and S. P. Voinigescu, "A 1GHz bandwidth low-pass ΔΣ ADC with 20GHz to 50GHz adjustable sampling rate," *IEEE Radio Frequency Integrated Circuits Symposium*, vol. 2, 2008, pp. 181–184.

[9] F. G. Stremler, *Communication systems*, New York: Pearson Education, 2005.

[10] L. Vannuccini, J. E. Earis, P. Helistö, B. M. G. Cheetham, M. Rossi, A. R. A. Sovijärvi and J. Vanderschoot, "Capturing and preprocessing of respiratory sounds," *Eur Respir Rev*, vol. 10, 2000, pp. 616–620.
[11] G. C. Chang and Y. F. Lai, "Performance evaluation and enhancement of lung sound recognition system in two real noisy environments," *Computer Methods and Programs in Biomedicine*, vol. 97, 2010, pp. 141-150.
[12] N. Gavriely, Y. Palti, G. Alroy and J. B. Grotberg. "Measurement and theory of wheezing breath sounds," *J Appl Physiol*, vol 57, 1984, pp. 481–492.
[13] H. Pasterkamp, A. Tal, F. Leahy, R. Fenton and V. Chernick. "The effect of anticholinergic treatment on postexertional wheezing in asthma studied by phonopneumography and spirometry," *Am Rev Respir Dis*, vol. 132, 1985, pp. 16–21
[14] B. Y. Lu, H. D. Wu, S. R. Shih, F. C. Chong, M. L. Hsueh and Y. L. Chen , "Combination of frequency and amplitude modulated Model for the synthesis of normal and wheezing sounds," *Austra Phy Eng Sci Med*, vol.34, 2011, pp. 449-457.
[15] V. M. Rossi, A. R. A. Sovijärvi, P. Piirilä, L. Vannuccini, F. Dalmasso and J. Vanderschoot, "Environmental and subject conditions and breathing manoeuvres for respiratory sound recording," *Eur Respir Rev*, vol. 10, 2000, pp. 611–615.
[16] PICO 900 Series Scope. Available: http://www.picotech.com/picoscope9000.html
[17] T. Wheel, *Electronic communication systems for technicians*, New York: Pearson Education, 2006.
[18] *User manual*, 3M Littman electronic stethoscope, Model 3200.

Bing-Yuh Benjamin Lu received his BS degree in electrical engineering from National Central University and his MS and PhD degrees in electrical engineering from National Taiwan University in 1988, 1993, and 2000, respectively.
He was an instructor at the Department of Electronic Engineering, Tungnan University, New Taipei City, Taiwan, Republic of China in 1992. In 2000, he became an associate professor in the same department. He joined the Department of Information Management, St. Mary's Medicine, Nursing, and Management College, Yi-Lan, Taiwan, Republic of China in 2011, and served as an associate professor and the department head in the duration of 2011 to 2013. He is currently an associate professor in Department of Electronic Engineering, Tungnan University, New Taipei City, Taiwan, Republic of China. He is interested in computer simulation, medical engineering, acoustics, lung sound and educational researches. He is a member of IEEE.
His academic interests in focus on acoustics, medical engineering, computer simulation, educational applications of engineering, and pulmonary signal processing especially about lung sound and pulmonary mechanism.

Liman Tseng is a lecturer in the Nursing Department, St. Mary's Junior College of Medicine, Nursing and Management, Yi-Lan, Taiwan. She has a MS degree in nursing, Tzu Chi University in 2005, who attended a PhD degree in public health, in Tzu Chi University in 2009. Her research and teaching interests include nursing administration, surgical nursing, oncology nursing and pediatrics nursing.

Ling-Yuan Hsu was born in I-Lan, Taiwan, R.O.C., in 1975. He received the M.S. degree in computer science and information engineering from National Dong Hwa University, Hualien, Taiwan, in 2004, and Ph.D. degree in computer science and information engineering at National Taiwan University of Science and Technology, Taipei, Taiwan, in 2013. Currently, he is an assistant professor in the Department of Information Management of ST. Mary's Medicine, Nursing, and Management College. His research interests include artificial intelligence, image processing, evolutionary computation, and computer networks.

Shr-Hong Tang was born in Taiwan in 1978. He received Ph.D. in Engineering from National Chiao Tung University (Hsin-Chu, Taiwan) in 2007. His major field of study was applied information technology in engineering. He is currently an assistant profession in St. Mary's Medicine, Nursing and Management College, and pursuing an advanced degree in computer science from National Taipei University of Education (Taipei, Taiwan)

Yu-Lin Weng received her M.S. degree in Information Management from National Taichung University of Science and Technology and her MS degrees in Network Learning technology from National Central University in 2006 and 2008, respectively. Currently, she is pursuing a PhD degree in Information Technology at National Taiwan University of Science and Technology .She was an instructor at the Department of Information Technology, St. Mary's Medicine, Nursing and Management College, Yi-Lan, Taiwan, Republic of China from 2008. She is currently the head of the Registration Section, St. Mary's Medicine, Nursing, and Management College. She is interested in digital learning, adaptive testing, and database systems.

Man-Na Hsueh received her MS degree in nursing from National Taipei University of Nursing and Health Sciences in 2004. Over the past 35 years, she has been teaching at St. Mary's Nursing School (before 2005) and St. Mary's Medicine, Nursing, and Management (after 2005). She was the Director of the Department of Students' Affairs from 1991 to

1997, the Director of the Department of Students' Internship Affairs from 2004 to 2005, and the Representative President and Director of the Department of Academic Affairs in 2005. She is currently the head of Department of Nursing at St. Mary's Junior College of Medicine, Nursing, and Management . She is interested in elder person's health care and in community medicine.

Mei-Ju Su received Ph.D. degree at Graduate Institute of Electronics Engineering, National Taiwan University since 2004 to 2010, Master degree at Department of Electrical Engineering, Syracuse University since 1996 tp 1997. Now, her researches focus on biosignal processing, medical device, tele-healthcare system, and sensor design. She had been worked at Industry for 8 years since 1998~2005 on network IC design and SiS Company, medical instrument system design at Biomedical Technology and Device Research Laboratories of Industrial Technology Research Institute. Now she is the associate professor of department of Biomedical engineering, Yuanpei University.

Huey-Dong Wu was graduated with a M.D. degree from Medicine Department of National Taiwan University in 1984. He received resident training in National Taiwan University Hospital (NTUH) and trained as visiting scholar in ULCA later. He service in the Department of Internal Medicine and Department of Integrated Diagnostic & Therapeutics in National Taiwan University Hospital (NTUH) till now. He is the chief of the division of Respiratory Care, Department of Integrated Diagnostic & Therapeutics in NTUH. His academic interests in focus on respiratory care, pulmonary pathophysiology and pulmonary signal processing especially about lung sound and pulmonary mechanism.

Jin-Shin Lai was born in Taipei, Taiwan, Republic of China (R.O.C.) in 1949. He received the M.D. degree from the Medical School of National Taiwan University, Taipei, Taiwan in 1974. He completed residency training in the Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital, Taiwan in 1978.He was a Lecturer, Associate Professor, and Professor of Medical School, College of Medicine, National Taiwan University in 1980, 1984 and 1996, respectively.

Now, he is the Director of Health Science and Wellness Center, National Taiwan University and Director of Sports Medicine Center, National Taiwan University Hospital. His professional interests include medical informatics, sports medicine, rehabilitation medicine, biomechanics, and rehabilitation engineering. He and his research group had published a series of studies about Physiological Effects of Tai Chi Chuan Training.

Prof. Lai has been the Chairman of School of Rehabilitation Medicine (1988-1992), Chairman of School of Occupational Therapy (1992-1998), Chairman of Department of Physical Medicine & Rehabilitation (1993-1999), Chief of Division of Medical Informatics (2001-2007), College of Medicine, National Taiwan University, respectively. He also served as the Chairman of Department of Information Technology & Service, National Taiwan University Hospital (1998-2004), and elected as the President of Rehabilitation Medicine Association of Taiwan (R.O.C.) 1993-1999, the President of Sports Medicine Association of Taiwan (R.O.C.) 1997-2001, the Chairman of Scientific Commission of Asian Federation of Sports Medicine (AFSM) 1990-2000, the Treasurer of Asian Federation of Sports Medicine (AFSM) 2001-2004 and the Chair of HL-7 Taiwan (2005-2009).

Techniques for System of Systems Engineering in Construction of a Smart Tourism Industry Information System

Liming Bai*

*College of Business Administration, Guangxi University of Finance and Economics,100 Mingxiu West Road, Xixiangtang District, Nanning, Guangxi Zhuang Autonomous Region 530003, P.R.CHINA

putaoren2000@aliyun.com

Abstract—**Currently, as many disciplines begin to cultivate a set of core methodologies, system of systems engineering (SoSE), which has its root in context of military programming, becomes a significant research focus and provides a new perspective to solve the emerging "system of systems" challenges in industrial analysis. Meanwhile, modern communication and information technologies have provided greater possibilities for socio-economic sectors to execute smarter decisions, and these technologies may advance the industrial application of SoSE. Therefore, targeting the rareness of government-oriented intelligent decision support system (DSS), and guided by the underlying system of systems thinking, this paper proposes a technical framework for designing a policy maker-responsive smart information system which focuses: (1) system of systems structural architecting; (2) geographical simulation using time-series remotely sensed data, GIS instrument and simulation bodies such as cellular automata (CA) and multi-agent systems (MAS); (3) SoS evolution description through network analysis and intelligent computing; (4) measurement of SoS effectiveness with two-tier and four-grade method; and (5) SoSE program for industrial optimization. Its application in tourism analysis will provide a smarter base for industrial policy-making, planning and forecasting, and will help reduce risk and cost in industrial restructuring. For this relatively new field of SoSE application, tools and methods are not perfect, so it is important to draw together academia, government, industrial organizations and enterprises to collaborate for further valuable achievement.**

Keyword—**System of systems engineering (SoSE), geographical simulation, tourism, smart industry information system, intelligent computing**

I. INTRODUCTION

IN the information age, as computer communication, internet of things, distributed control, cloud computing and many other fields of technology advance, originally separated individual systems (or system elements) become capable of connecting to each other by information "ties", forming a meta-system that involves various heterogeneous distributed systems (including policies, economies or technologies) [1]. Decision makers within government and industry encounter a prevalent problem of increasing complexity, which are of system-of-systems type [1]. This is particularly evident in tourism industry. Tourism plays an important role in national economies of many countries. Therefore, growth of the industry results in complex networks consisted of expanding conventional components and newly generated cross-sector systems. These growing complexities, which is of system of systems type, have challenged the system-engineering-principles based industrial analysis in investigating the industry architecture and evolution, making effective decision-support become unmanageable [1].

This becomes a problem to be solved in construction of decision support system (DSS) as well. There are 3 different types of tourism-related information systems: (1) tourist-oriented information system (TIS). It provides information on tourist attractions, itineraries, logging and accommodations and tour guide, ect. for tourists or potential tourists. For instance, Ren developed a Tourism Information System for Wuhan City, China based on MapGIS K9, with the aim, principle, overall framework and function module design. The study emphasized the feasibility and broad prospect of the GIS technology in construction of such systems [2]; (2) Tourism enterprise-oriented information systems (TES). It is a tourism management information system aids managerial or marketing behaviors of individual corporate. Taking the Sanzhualun National Forest Park as an example, Wang and Chen introduced the use of WebGIS technology in forest park tourism information system development [3]. Wan and Tang established a persistent data frame using .NET platform and Castle technologies, and applied it into developing a travel management system [4]. (3) Policy/industrial decision maker-oriented system (Tourism industry information system, TIIS). It mainly provides support for decision making process of governments, industrial organization. Li and Zhan presented the realizing schemes and process of the object-oriented data model in the construction of a Tourism Information

Manuscript received December 31, 2012. This work was supported in part by the National Natural Science Foundation of China under Grant 71163004, the Humanities and Social Sciences Foundation of Guangxi under Grant SK 13LX363 and the Research Foundation of Guangxi University of Finance and Economics under Grant 2012C17.

L. M. Bai is with the Guangxi University of Finance and Economics, Nanning, Guangxi Zhuang Autonomous Region 530003, China (phone: 139-7868-4395; fax: 077-1384-3160; e-mail: putaoren2000@ aliyun.com).

Management System (TIMS) [5]. A set of synthetical function indexes were incorporated into the system so that it can provide city tourism administration agents several decision making services such as tourist attraction analysis, regional tourism planning, tourism product and tourist itinerary design.

The information provided by the 3 types of information systems is significantly different. Differs from TIS and TES, TIIS has to concern more on macro environment dynamics and complex stakeholder networks, and it has more work to with big data. Therefore, smart or virtual communication is widely used in micro-management practices (e.g., corporate or profession administration), but when it comes to macro-management on industrial or regional level, due to the need of both large-scale geographical information and diverse simulation models, government/industry-oriented intelligent decision support system is relatively rare.

Therefore, inspired by the embedded "system of systems (SoS) thinking" in the text of military programming, this paper attempts to architect a policy maker-responsive smart tourism industry information system and proposes the related modeling techniques and approaches.

II. System of Systems and System of Systems Engineering

A. System of Systems

Development of information and communication technologies have changed the form and frequency of the interactions among social systems, therefore, at the beginning of the 21st century, the concept of "system of systems" [6-7] began to become an emerging internationally research focus in system science[8]. Many research centers, such as the National Centers for Systems of Systems Engineering (NCSoSE) and the SoS Engineering Center of Excellence (SoSECE), were established to study complex systems of systems problem domains in different fields.

Though without a widely accepted definition, the core of concept is commonly recognized [9]. Sage and Cuppan suggested that a SoS has to meet the Maier's five criteria [10], and Keating defined SoS as a meta-system consists of multiple individual systems that diversified in operating context, geographic location, conceptual architecture and function mode [11]. To some extent, SoS is a task-oriented system alliance, which possess the resources and capabilities of each component system and performs more productive than the sum of the individual systems. Therefore, SoS is often designed to solve the multi-system interaction and /or integration problems[12], dealing with inherent complexity domains [13].

In recent years, differences between SoS and general system have been recognized, and architectural design, resource confrontation and performance evaluation for SoS have become increasingly emphasized in national defense, aerospace exploration, information or communication networks, transportation, energy, healthcare, environment preservation, management and many other application fields [12].

B. Socio-Technical Features of Systems of Systems

Differentiated from simple general system, system of systems has several significant socio-technical features [8]: (1) large in scale, complex in architecture, and composed of multiple component systems; (2) the component systems demonstrate wide geographic distribution, and they have operational, functional or managerial independence; (3) having unfixed targets under a specified purpose, and the SoS resources can be dynamically configured to meet diversified needs of multiple tasks; (4) components systems share interdependence, interoperability and can run concurrently in executing a specified plan; (5) developing under centralized planning and administration, and producing behavioral or functional emergence in process of evolution; (6) it is critical to explore and coordinate the capabilities/resources of different component systems or stakeholders to gain collectively provided upgraded capability.

C. System of Systems Engineering

Definition, abstraction, modeling and analysis of system of systems challenges ascribe to system of systems engineering (SoSE), which is a set of developing processes, instruments, and approaches providing design or re-construction solutions.

System of systems engineering methodology is initialized by U.S. Department of Defense, and is developed mostly in military applications in many countries. But in recent years, its superiority in dealing with emerging complex socio-economic or technical problems is increasingly recognized, driving the non-defense applications. SoSE is not simply systems engineering of monolithic and complex systems, as the solution design emphasizes uncertainties in the mission requirements and the component systems, and it contains multi-level and multi-dimensional consistent engineering programs [14-15]. Rather than systems engineering concentrating on building the system right, SoSE focuses on configuration of selected systems and their resources and interactions to meet the requirements of SoS capability.

There is not a single unified consensus for processes involved in system of systems engineering. DeLaurentis suggested a three-phase SoSE framework where a SoS problem is defined, modeled and analyzed [16]. Zhang et al proposed a two-level and four-grade measure approach focusing on SoS performance [12].

III. Technologies for Smart System Architecting

In recent years, systems science (especially complexity) and geo-spatial information science are emphasized in tourism study. complex networks analysis [17], cellular automata (CA) and multi-agent systems (MAS) [18] is widely used.

However, many studies only emphasized the static industrial architectures, but ignored the dynamic evolution mechanism towards the expected goal [19]. Yang analyzed the tourism industry network and its evolution, and pointed out that the formation and evolution of the networks were driven by internal heterogeneity and external uncertainty [19]. Xue and Weng analyzed regional tourism spatial layout using economics deductive model combined with agent-based calculation. The microscopic background, influence factors and dynamic process of tourism spatial architecture changes could be "dynamically" observed through the interaction of a large number of micro agents (enterprises, consumers etc.)[20].

These research applications provide a meaningful enlightment for constructing a smart tourism industry information system. Guided by the underlying system of systems engineering thinking, and based on an integration of information and communication technologies, this paper proposes a technical framework for designing a smart information system. The phases include: (1) system of systems structural architecting; (2) geographical simulation using remotely sensed data, GIS instrument and intelligent computational agents; (3) SoS evolution description through network analysis and intelligent computing; (4) measurement of SoS effectiveness with two-tier and four-grade measure method; and (5) SoSE program for industrial optimization.

The technical route is shown in Fig. 1, and the detailed information will be stated in the subsequent sections.

A. System of Systems Structural Architecting

Tourism industrial SoS is a complex network of geographically distributed resources, market entities, consumers, stakeholders and their intensive communications, which evolutes both spatially and temporally in architecture and component systems. Emergence may be produced in the dynamic SoS evolution. Therefore, the first phase for constructing the smart industry information system is SoS architecting, namely, properly converting the industrial SoS into virtual properties in the simulation system. The structural architecting for tourism SoS is demonstrated in Fig. 2.

The basic elements of the tourism SoS to be simulated include resource entity, SoS architecture, SoS process and environment, and the component system itself is a composite element. Primary task of the study is to identify the basic units and the consequently assembled component systems, as well as structural relationships between the tourism SoS and the component systems. The smart information system consists of tourist geographical system, landscape system, industrial layout system and social System.

Fig. 2 only provides a simple conceptual model, and further improvement should be made to elaborate more elements, component systems and the embedded hierarchy and interaction networks. Moreover, an enhanced functional architecting is in need to identify the requirements for enhanced SoS functions.

B. Geographical Simulation: Cellular Automata with Geographical Multi-agent

In the second phase, the tourism SoS will be geographically simulated based on remotely sensed data, GIS modeling platform, CA and MAS instruments. Time-series simulation outputs though knowledge discovery (KDD) and data mining can provide support for the following analysis missions.

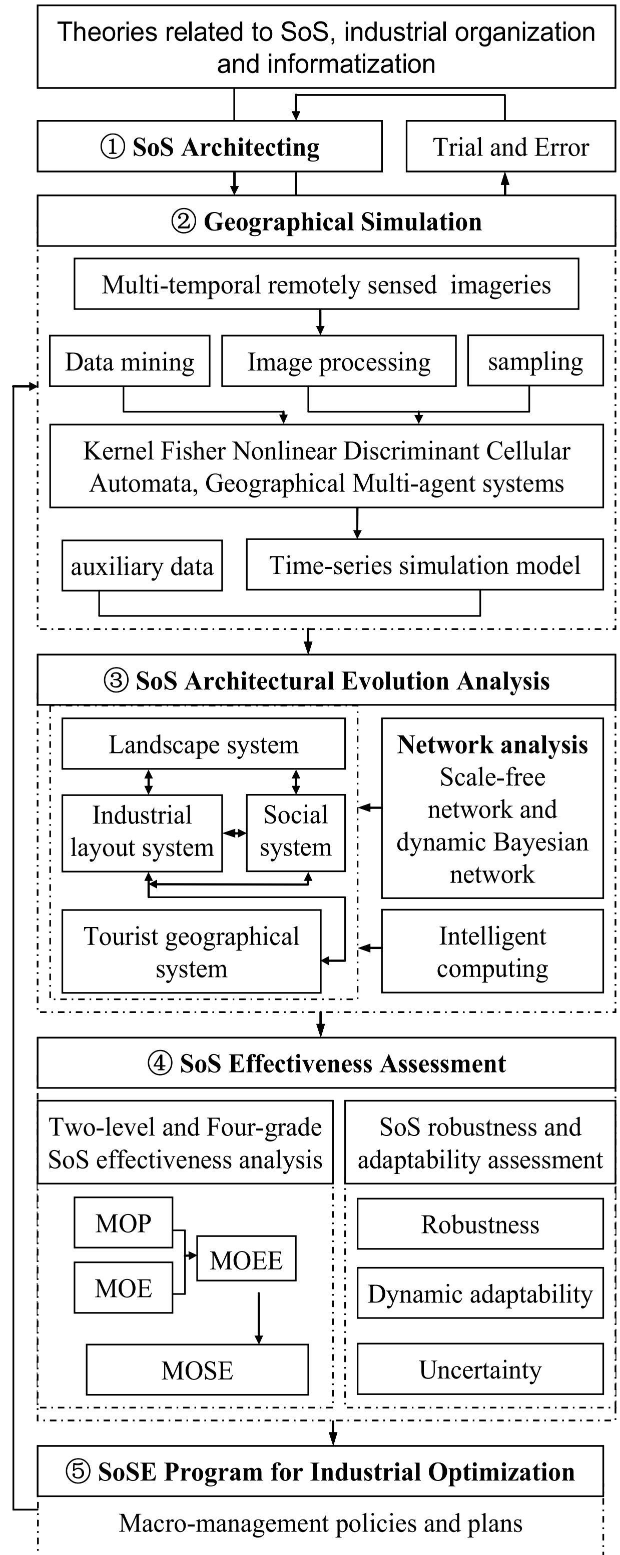


Fig. 1. Technical route for constructing the smart tourism industrial information system. The designation focuses 5 critical subsequent processes: 1) system of systems structural architecting; 2) geographical simulation using time-series remotely sensed data, GIS instrument and simulation bodies such as cellular automata (CA) and multi-agent systems (MAS); 3) SoS evolution description through network analysis and intelligent computing; 4) measurement of SoS effectiveness through two-tier and four-grade measure method; and 5) SoSE program for industrial optimization. Each of these processes is capable of outputting relevant industrial information.

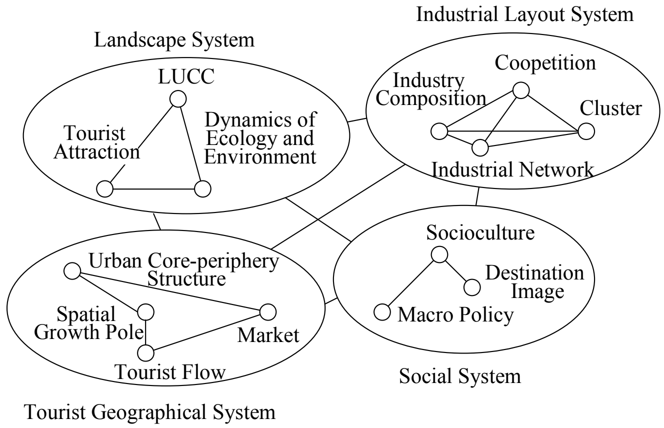


Fig. 2. Conceptual model for the smart tourism industrial information system.

The core of geographical simulation is to establish scientific geospatial systems model using appropriate research methodologies, based on application of complexity theory and combined with the inherent law of geography. Cellular automata (CA) and multi-agent system (MAS) are suited to study the complex geographic phenomenon [17].

CA is a grid dynamics model that is discrete in time, space and state, and is local in spatial interaction and temporal causal relationship. It is an important research tool and theoretical method branch of artificial life. Many studies have shown, the kernel Fisher nonlinear discriminant learning machine is superior to other methods in automatically extracting the CA transition rules. However, CA cell can not move, having limitations in dynamic environment and cell interactions, thus introduction of MAS modeling proves useful. MAS stems from a complex adaptive system, it is composed of multiple computing units (agent) mutually interacting. Spatial MAS resolves the macro space pattern formation though the interaction among microscopic individual (cell) or agent, it is very close to the real geographical world, and can reflect the characteristics of complex space systems in terms of emergence, chaos and evolution, thus has more analog advantage. MAS is a SoSE method, which can ensure the coherence of SoSE applications in tourism geographical simulation.

C. SoS Evolution Description: Network Analysis and Intelligent Computing

System network approaches such as small-world network analysis, scale-free network analysis, and social influence network theory (SINT) with dynamic Bayesian network, can be employed to simulate the growth and evolution of the tourism industry SoS. They are also effective in describing the complex interactions between the component systems. In this process, intelligent computing will play a critical role in data mining.

Small-world Network Analysis

A network in the real world usually has both certain rules and some randomness. To describe the transition from a network of rules to a random one, Watts and Strogatz introduced an interesting WS Small-world Network Model in 1998 [21]. The model can be used for analog of the derivative and evolution of a newborn or a simple component system in the tourism SoS.

Scale-free Network Analysis

Unlike the degree distribution for the exponent network, that of many large-scale complex network is a descending curve, following a power law distribution $P(k) \sim k^{-\gamma}$. As explanation of this power-law distribution, Barabási and Albert proposed the Scale-free Network Model [22, 23]. The evolution of BA scale-free network mainly has two aspects, namely growth and preferential attachment. New nodes are constantly added into the network, and it is with high probability for the new nodes to connect to the existing node with larger number of connections. The generation algorithm is as follows:

--Growth: Initially there are m_0 nodes in the network, and one new node is added at each step, the node is connected to one of the existing $m \le m_0$ nodes.

--Preferential attachment: When selecting the node to which the new node attached in the network, choice is made according to Π probability. If the degree of the node i is k_i, then the probability of the existing node to be selected is defined as:

$$\prod(k_i) = \frac{k_i}{\sum_j k_j} \tag{1}$$

In this way, after t steps, a network with $N = t + m_0$ nodes and mt edges is generated. In such a network, there are a few nodes having a very high degree (i.e., the number of nodes connected to it), while the majority of the nodes show relatively low degree. Tourism SoS is open, elements (nodes) are constantly added and the attachment features as unequal probability. Therefore, when the tourism SoS and its component systems have evolved into the near mature or mature stage of the life cycle, the application of scale-free network model will produce more realistic descriptions.

Social Influence Network and Dynamic Bayesian networks

To achieve the desired objective for the smart tourism industry information system, the SoSE program guiding the SoS evolution should make clear how to configure the SoS and achieve the best SoS configuration through circulated SoS measurements. The Social Influence Network Theory (SINT), combined with Dynamic Bayesian Networks (DBN), can fulfill this dynamic process. The SINT depends on the individual's coordination processing on the conflict viewpoints, stances and attitudes. But the output of the conflict processing results depends on the structural relationship between the individuals, such post, influence of interaction and one's sensibility (that is, to what extent can an individual be influenced) to it. Its expression is as follows:

$$y_i^{(t+1)} = a_i(w_{i1}y_1^{(t)} + w_{i2}y_2^{(t)} + \cdots + w_{iN}y_N^{(t)}) + (1-a_i)y_i^{(1)} \tag{2}$$

Wherein, t is period of time; y represents the viewpoint, position or attitude of the individual i ($i = 1, 2, \cdots, N$, N is the number of individuals in the group) in time period t, its initial value is $y_i^{(1)}$; w_{ij} is the weight for the influence of individual i upon individual j ($0 \le w_{ij} \le 1$), and $\sum_{j=1}^{N} w_{ij} = 1$; individual i's sensibility $a_i = \sum_{j=1, j \ne i}^{N} w_{ij}$

One of the methods to estimate w_{ij} is DBN. Bayesian network is a product of the combination of probability theory and graph theory, and is composed of network structure and conditional probability distribution. Its specific applicative architecture is determined based on knowledge principle and the structure of variable relationship. For a Bayesian network with total number of N nodes, its network architecture depends on the following set of conditional independence assumptions:

$$P(v_i | v_1, \cdots v_{i-1}, v_{i+1}, \cdots v_N) = p(v_i | parent(v_i)) \quad (3)$$

Wherein, $i = 1, 2, \cdots, N$; $parent(v_i)$ is the parent node corresponding to node v_i. Thus, it can be inferred from the chain rule of probability that, the joint probability distribution for the node set V is:

$$P(V) = \prod_{i=1}^{N} p(v_i | v_1, \cdots v_{i-1}, v_{i+1}, \cdots v_N) = \prod_{i=1}^{N} p(v_i | parent(v_i)) \quad (4)$$

DBN provides a strong support for the extraction of the complex interactions between systems or elements within the tourism SoS. Therefore, it can be used in overall analysis of the SoS, producing in the smart tourism industry information system a resource allocation program conducive to industrial optimization.

Intelligent Computing in SoSE

Application of intelligent computing may contribute to improving the system of systems engineering effectiveness in the smart industrial system's operation. As an essential stream of artificial intelligence, intelligent computing is an empirical computer thinking program in data processing after sequential set of steps of intelligent acquisition and recognition. The computing model is of hierarchical nature, and it generally features six levels[24], namely, operational simulation, existential experience, evaluation, inference of change limit, intelligent computing experience system and multi-link simulation and behavior capture (Fig. 3).

Intelligent computing body is capable of self-growth and being more adaptive with time and record of changes. Algorithms for intelligent computing includes neural networks, machine learning, bio-computing, fuzzy logic, pattern etc.

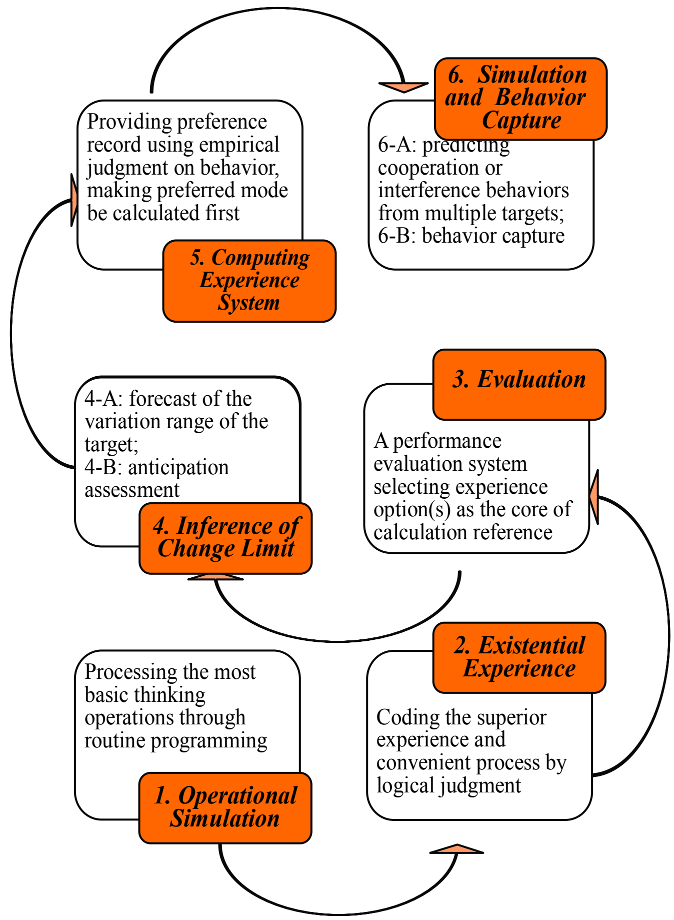


Fig. 3. Intelligent computing process in system of systems engineering. Six levels are included with respect to: operational simulation, existential experience, evaluation, inference of change limit, intelligent computing experience system and multi-link simulation and behavior capture.

Intelligent computing frame and algorithms can be used in system of systems engineering processes including data mining, complex network analysis, big data processing and policy forecast. The effectiveness and functionality of the smart tourism information system will be significantly improved through hybrid embedded intelligent computing algorithms and systems.

D. Emergence and SoS Effectiveness Measurement: Two-level and Four-grade Method

Measurement and evaluation of a SoS can be conducted on aspects including SoS effectiveness, robustness and adaptability. The overall performance is the extent of to which the SoS have a beneficial role in the mission process and outcome. Corresponds to hierarchical characteristics of SoS architecture, a two-level and four-grade measure method was proposed in [12:253-259]. The method mainly includes measures of performance (MOP), measures of effectiveness (MOE), measures of emergence effectiveness (MOEE) and measures of system of systems effectiveness (MOSE). Among the four grades, MOP and MOE are oriented to component subsystems of a SoS, while MOEE, and MOSE are oriented to the SoS' top-level mission operations.

Measures of Performance (MOP)

Performance of system unit is the inherent properties or characteristics of a given SoS. Suppose a SOS consisted of n system unit, namely

$$SoS = \{s_1, s_2, \cdots, s_n\} \tag{5}$$

Wherein, p_i is the subsystem level performance corresponding to any system unit s_i $(i = 1,2,\cdots,n)$. Suppose the numer of performance metrics for system unit s_i was r_i, p_i will be presented as:

$$p_i = \{p_{i,1}, p_{i,2}, \cdots, p_{i,r_i}\} \tag{6}$$

Performance metrics of system unit s_i would be subject to the restriction of a specific low-end performance threshold (denoted by p_i^*) as well as the technological constraints for high-end performance.

Measures of Effectiveness (MOE)

Effectiveness aggregation of multiple cooperating and mutually administrative system units constitute input into measures of effectiveness for emergence-level behavior characteristics. MOE for each system unit is likely to contain multiple aspects, i.e. for any system unit s_i $(i = 1,2,\cdots,n)$, its corresponding system effectiveness will be:

$$e_i = \{e_{i,1}, e_{i,2}, \cdots, e_{i,l_i}\} \tag{7}$$

Wherein, l_i denotes the dimension of system effectiveness e_i for system unit s_i; for each specific system effectiveness dimension, its corresponding MOE is:

$$e_{i,j} = f_{i,j}(p_i) = f_{i,j}(p_{i,1}, p_{i,2}, \cdots, p_{i,r_i}), \forall j \le l_i \tag{8}$$

Similar to MOP, there is a minimum threshold $e_{i,j}^*$ for MOE as well.

Measures of Emergence Effectiveness (MOEE)

A SoS comprises a number of hierarchical systems, each system composite has a corresponding emergence grade. while the overall characteristics and behavior of the SoS or its system composites are an outcome of emergence of all cooperative and interconnected subsystems within them. Therefore, MOEE serves as the core of SoS measurement, and also shows certain hierarchical characteristics.

--MOEE on bottom-line emergence grade: system composites corresponding to the bottom-line emergence grade are system units. Suppose MOEE for a bottom-line emergence grade was EE_i, and it contained o_i aspects, where each aspect corresponded to a sub-operation of the mission, that is:

$$EE_i = \{EE_{i,1}, EE_{i,2}, \cdots, EE_{i,o_i}\} \tag{9}$$

Therefore,

$$EE_i = \Omega_i(m_i, e_k, e_p, e_q, \cdots) \tag{10}$$

Where Ω_i is MOEE function corresponding to emergence grade i; e_k, e_p, e_q indicates the effectiveness values of the interrelated system units s_k, s_p, s_q, which act in accordance with homologous emergence levels and complete missions corresponding to SoS or system composite, while m_i indicates the quantitative architecture of these interrelated system units

--MOEE on non-bottom-line emergence grade: system composite on any non-bottom-line emergence grade contains at least one lower-leveled system complex. They share the same measures and constraint function types with the bottom-line emergence grade MOEE, just differing on each variable's indicating meaning.

Measures of System of Systems Effectiveness (MOSE)

Measures of system of systems effectiveness refers to assessment of SoS goal's final achievement status, determined by the effectiveness of the behavioral characteristics on global emergence grade. Suppose SoS' overall MOSE was E, and it contained l aspects, namely

$$E = \{E_1, E_2, \cdots, E_l\} \tag{11}$$

Then,

$$E = E(N, EE) = E(N, EE_1, EE_2, \cdots, EE_o) \tag{12}$$

Wherein, E denotes function of MOSE for SoS; EE is effectiveness function of behavior characteristics on homologous emergence grade, and it includes o aspects, corresponding to o sub-operations of the SoS respectively, namely $EE = \{EE_1, EE_2, \cdots, EE_o\}$; While N indicates the quantitative structure of these sub-operations.

Eventually, industrial optimization projects can be carried out based on the "hidden order" of the tourism SoS evolution that discovered by the smart industrial system. It is the ultimate aim of constructing such a system.

IV. CONCLUSION

Currently, as many disciplines begin to cultivate a set of core methodologies [25], SoSE, which has its root in context of military programming, is proven to be a notable research focus, providing a new perspective to solve the emerging system of systems challenges in industrial analysis. SoS thinking is of great significance in cross-disciplinary study [12], and it may lead a new research paradigm. Its application in tourism analysis will provide a solid foundation for industrial policy-making, planning and forecasting, and will help reduce risk and cost in industrial restructuring.

Modern communication and information technologies provide greater possibilities for socio-economic sectors to execute smarter decisions. These technologies will advance the SoSE industrial application. Therefore, targeting the rareness of government oriented intelligent decision support system (DSS), and inspired by the "system of systems thinking", this paper presents a technical route for constructing a policy maker-responsive smart tourism industry information system.

System of systems architecting, geographical simulation, network analysis with intelligent computing and SoS effectiveness evaluation are focused in the smart system construction. The overall design include: 1) system of systems structural architecting; 2) geographical simulation using geo-science instruments, CA and MAS; 3) SoS evolution description through network analysis and intelligent computing; 4) SoS effectiveness measurement using two-tier and four-grade method; and 5) SoSE program for industrial optimization.

However, SoS tools and methods are not perfect, so it is important to draw together academia, government, industrial organizations and enterprises to collaborate in the related studies. According to the 10 hot issues leading the future development of SoSE [7, 8], future research directions for the tourism SoSE may include: (1) Flexibility, adaptability, and capability of rapid recovery; (2) Model-driven SoS architecture; (3) Multi-view product for SoS architecture; (4) Net-centric vulnerability; (5) Evolution; (6) Guided emergence and capability engineering.

References

[1] D. DeLaurentis and R. K. Callaway, "A system-of-systems perspective for public policy decisions," *Rev. Policy Res.*, vol. 21, pp. 829-837, November 2004.

[2] Z. Ren, "Design and development of Wuhan municipal tourism information system based on MapGIS K9," *Sci. Tech. Manage. Land Resour.*, Vol. 28, pp. 109-114, October 2011. (In Chinese)

[3] J. D. Wang and X. Y. Chen, "WebGIS-based national forest park tourism information system," *Jiangxi Forest. Technol.*, vol. 6, pp.35-37 ,2011. (In Chinese)

[4] C. P. Wan and H. J. Tang, "Design of travel management system based on ASP.NET and Castle framework," *J. Chengdu Univ. Inf. Technol.*, vol.22, pp. 458-461, August 2007. (In Chinese)

[5] M. Li and C. W. Zhan, "Research on the application of object-oriented data model technologies in the tourism information management," *J. Chongqing Inst. of Tech.*, Vol.20, pp. 115-118, 125, May 2006. (In Chinese)

[6] M. W. Maier, "Architecting principles for Systems-of-Systems," *Syst. Eng.*, Vol. 4, pp. 267-84, 1998.

[7] A. Sousa, P. S. Kovacic, and C. Keating, "System of systems engineering: an emerging multidiscipline," *Int. J. Syst. Syst. Eng.*, vol. 1, pp.1-1 7, 2008.

[8] Y. J. Tan and Q. S. Zhao, "A research on system-of-systems and system-of-systems engineering," *J. CAEIT*, vol. 6, pp. 441-445, October 2011.

[9] Mane, M., W. A. Crossley, and Nusawardhana, "System of systems inspired aircraft sizing and airline resource allocation via decomposition," *J. Aircraft*, Vol. 44, pp. 1222-1235, July-August 2007.

[10] A. P. Sage and C. D. Cuppan, "On the systems engineering and management of systems of systems and federations of systems," *Inf., Knowl., Syst. Manage.*, Vol. 2, 2001, pp. 325-45.

[11] C. Keating, R. Rogers, R. Unal, D. Dryer, A. Sousa-Poza, R. Safford, W. Peterson, and G. Rabadi, "System of systems engineering," *Eng. Manage. J.*, Vol. 15, pp. 36-45, September 2003.

[12] W. M. Zhang, Z. Liu, and D. S. Yang, et al., *Theory and methodology for system of systems engineering*. Beijing, China: Science Press, 2010.

[13] X. F. Hu, B. Zhang, "SoS complexity and SoS engineering", *J. CAEIT*, vol. 6, pp. 446-450, 2011.

[14] D. DeLaurentis, "Understanding Transportation as a System-of-Systems Design Problem," 43rd AIAA Aero. Sci. Meeting, Reno, Nevada, January 10–13, 2005. AIAA-2005-0123.

[15] D. DeLaurentis, O. Sindiy, and W. Stein, "Developing Sustainable Space Exploration via a System-of-Systems Approach," AIAA Space 2006 Conf., San Jose, CA

[16] X. H. Bian. *The forming mechanism of urban tourism spatial structure: A case study of the Yangtze River Delta*, Shanghai, China: Gezhi Press, 2008

[17] J. P. Liu, Z. P. Shen, and C. Mo. "The application of complex networks theory in tourism industry cluster of Shanghai World Expo", *Econo. Geog.*, vol. 29, pp. (12):2023-2027, 2009.

[18] X. Li, J. A. Ye, X. P. Liu, and Q. S. Yang, *Geographical simulation systems: CA and MAS*, Beijing, China: Science Press, 2007.

[19] Y. Yang, D. L. Zhuang, and T. F. Tao, "The mechanism and the evolution of tourism industrial networks", *Ecol. Econo.*, vol. 7, pp. 91-94, 2010.

[20] L. Xue and J. Weng, "Study on microcosmic mechanism and dynamic simulation of China's regional tourism spatial structure evolution", *Tourism Tribune*, vol. 25, pp. 26-33, 2010.

[21] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small world' networks", *Nat.*, vol. 393, pp. 440-442, 1998.

[22] A. L. Barabási and R. Albert, "Emergence of scaling in random networks", *Sci.*, vol. 286, pp. 509-512, 1999.

[23] A. L. Barabási, R. Albert, and H. Jeong, "Mean-field theory for scale-free random networks", *Physica A*, vol. 272, pp.173-187, 1999.

[24] Hangxuan Planning, "Remote intelligent computing systems of complete algorithm for ditch project earthwork balance", China, 2012SR063598, 14 Jun. 2012.

[25] Check Teck Foo, "Implementing Sun Tzu's Art of War, system of systems (SoS) thinking: Integrating pilot's F22 Raptor cockpit and the brain of CEO," *Chinese Manage. Stud.*, Vol. 3, pp. 178-186, 2009.

Liming Bai (M'13) (Inner Mongolian, China, 1980) was awarded a doctorate in agriculture at Center of Forestry Remote Sensing Information Engineering, Central South University of Forestry and Technology, Changsha, Hunan Province, China in 2011. Her major field of study is remote sensing.

She is a Lecturer in College of Business Administration, Guangxi University of Finance and Economics. Her current research interests are tourism system of systems engineering and smart geographical simulation with intelligent computing, remote sensing and geographic information system (GIS).

Volume 3 Issue 3, Jan 2014, ISSN: 2288-0003

ICACT-TACT JOURNAL

Global IT Research Institute

1713 Obelisk, 216 Seohyunno, Bundang-gu, Sungnam Kyunggi-do, Republic of Korea 463-824
Business Licence Number : 220-82-07506, Contact: secretariat@icact.org Tel: +82-70-4146-4991