www.zte.com.cn/magazine/English

ISSN 1673-5188
CODEN ZCTOAK

ZTE COMMUNICATIONS

An International ICT R&D Journal Sponsored by ZTE Corporation

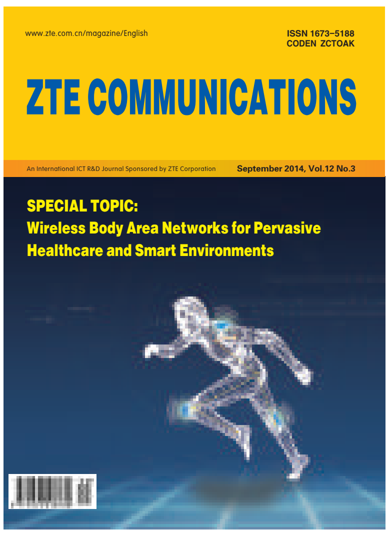
September 2014, Vol.12 No. 3

SPECIAL TOPIC:
Wireless Body Area Networks for Pervasive Healthcare and Smart Environments

ZTE中兴 ZTE COMMUNICATIONS

VOLUME 12 NUMBER 3 SEPTEMBER 2014

CONTENTS



Special Topic: Wireless Body Area Networks for Pervasive Healthcare and Smart Environments

CONTENTS

ZTE COMMUNICATIONS
Vol. 12 No. 3 (Issue 43)
Quarterly
First English Issue Published in 2003

Supervised by:
Anhui Science and Technology Department

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

Staff Members:
Editor-in-Chief: Sun Zhenge
Associate Editor-in-Chief: Zhao Jinming
Executive Associate Editor-in-Chief: Huang Xinming
Editor-in-Charge: Zhu Li
Editors: Paul Sleswick, Xu Ye, Yang Qinyi, Lu Dan
Producer: Yu Gang
Circulation Executive: Wang Pingping
Assistant: Wang Kun

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

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

Printed by:
Hefei Zhongjian Color Printing Company

Publication Date:
September 25, 2014

Publication Licenses:

ISSN 1673-5188
CN 34-1294/ TN

Advertising License:
皖合工商广字0058号

Annual Subscription:
RMB 80

Wireless Body Area Networks for Pervasive Healthcare and Smart Environments

Wireless body area networks (WBANs) use RF communication for interconnection of tiny sensor nodes located in, on, or in close proximity to the human body. A WBAN enables physiological signals, physical activity, and body position to be continuously monitored. Designing a WBAN is challenging because of the limited energy that a WBAN can consume and the limited processing capabilities of sensor nodes. Also, the radio communication environment is highly variable and prone to interference. Recent advances in wearable and implantable biosensors, short-range wireless communication, and low-power embedded processors are contributing to an increase in WBAN R&D aimed at addressing these issues. WBANs usually function as signal sources in larger, more intelligent systems used in applications that have the potential for great social and economic good. These larger systems are formed by connecting WBANs with external communication and computing infrastructure, e.g., cloud-computing services accessed through a smartphone that connects to the Internet via a wireless WAN. There is strong interest among researchers and medical practitioners in the development of intelligent systems based on WBAN. These systems enable pervasive e-healthcare applications, such as ambulatory monitoring of outpatients, as well as smarter environments that support context-aware applications, aspects of video gaming, monitoring of sports training regimes, and monitoring of emergency personnel and mission-critical workers. The purpose of this special issue is to survey WBAN in terms of state state-of-the-art technologies, latest developments, and useful applications. Original papers were solicited from experts on WBAN, and six of these papers were selected for peer-review and publication. Each paper covers a different aspect of WBAN.

The first paper, "Sensing, Signal Processing, and Communication for WBANs," by S. H. Fouladi, R. Chávez-Santiago, P. A. Floor, I. Balasingham, and Tor A. Ramstad, is a survey of recent research on signal processing related to sensor measurements in WBAN. The paper describes aspects of communication based on the IEEE 802.15.6 standard. The paper also describes state-of-the-art modeling for WBAN channels in all frequency bands specified in IEEE 802.15.6. The authors discuss the need for channel models for new frequency bands.

The second paper, "MAC Layer Resource Allocation for Wireless Body Area Networks," by Q. Shen, X. Shen, T. H. Luan, and J. Liu, describes a centralized MAC layer resource-allocation scheme for WBAN. The authors focus on mitigating interference between WBANs and reducing the amount of power consumed by sensors. This scheme involves a central controller that optimizes channel resource allocation according to channel and buffer state reported by smartphones. Temporal correlations of body area channels are exploited to minimize channel state reporting. A myopic policy is developed to solve the network design formulated as a partly observable optimization problem.

▶ Victor C. M. Leung

Victor C. M. Leung is a professor of electrical and computer engineering and TELUS Mobility Research Chair at the University of British Columbia, Canada. He researches wireless networks and mobile systems and has has co-authored more than 700 technical papers in journals and refereed conference proceedings. Several of these papers have won best-paper awards. Dr. Leung is a fellow of the IEEE, Royal Society of Canada, Canadian Academy of Engineering, and Engineering Institute of Canada. He has served on the editorial boards of *IEEE Journal on Selected Areas in Communications, IEEE Transactions on Computers, IEEE Transactions on Wireless Communications, IEEE Transactions on Vehicular Technology, Wireless Communications Letters*, and several other journals. He has led the technical program committees and organizing committees of numerous international conferences and workshops. Dr. Leung was awarded the APEBC Gold Medal in 1977. He received NSERC postgraduate scholarships from 1977 - 1981 and was awarded the IEEE Vancouver Section Centennial Award in 2011. He received the UBC Killam Research Prize in 2012.

▶ Hongke Zhang

Hongke Zhang received his MS degree and PhD degree in electrical and communication systems from the University of Electronic Science and Technology, China, in 1988 and 1992. From 1992 to 1994, he was a postdoctoral researcher at Beijing Jiaotong University. In July 1994, he began working in the School of Electronics and Information Engineering, Beijing Jiaotong University, and is now a professor in that school. He is the director of the National Engineering Lab for Next-Generation Internet Interconnection Devices, and her is the chief scientist on the Basic Research on Theories of Smart and Cooperative Networks project, which was founded by the National Basic Research Program of China ("973" Program). He is a member of the IEEE and chairman of IEEE 1888.2 Work Group.

Wireless Body Area Networks for Pervasive Healthcare and Smart Environments

Victor C. M. Leung and Hongke Zhang

The third paper, "Selective Cluster-Based Temperature Monitoring System for Homogeneous Wireless Sensor Networks," by S. Tyagi, S. Tanwar, S. K. Gupta, N. Kumar, and J. J. P. C. Rodrigues, describes a health monitoring system for critically ill patients as a case study for temperature-monitoring based on Enhanced LEACH Selective Cluster (E-LEACH-SC) routing protocol. E-LEACH-SC uses direct and selective cluster-based data transmission for short-range and long-range collection of data from ill patients. Simulations show that E-LEACH-SC significantly increases network lifetime compared to traditional LEACH and LEACH-SC protocols.

The fourth paper, "Prototype of Integrating Internet of Things and Emergency Service in an IP Multimedia Subsystem for Wireless Body Area Networks," by K.-D. Chang, J.-L. Chen, and H.-C. Chao, describes a common fabric for integrating the Internet of Things into the Internet and supporting emergency call processing so that critical WBAN data can be transferred. The paper describes a simulated bootstrap platform using 3GPP IP Multimedia Subsystem services as well as a prototype implementation. Experimental and simulation results show that the system is suitable for providing emergency services.

The fifth paper, "Smart Body Sensor Object Networking," by B. Khasnabish, describes the networking and internetworking of smart body sensor objects. The author proposes making body sensor objects smarter by giving them virtualization, predictive analytic, and proactive computing and communications capabilities. The author also describes use cases that include the relevant privacy and protocol requirements. General usage and deployment etiquette and relevant regulatory implications are also discussed.

The sixth paper, "E-Healthcare Supported by Big Data," by J. Liu, J. Wan, S. He, and Y. Zhang, describes how e-healthcare has increased transparency by making decades of stored health data searchable and usable. The authors give an overview of the architecture of e-healthcare, including four layers: data collection, data transport, data storage, and data analysis. Challenges in data security, data privacy, real-time delivery, and open standard interface are also discussed.

In closing, we would like to thank all the authors for their contributions and all the reviewers for their efforts in helping to improve the quality of the papers. We are grateful to the editorial office of *ZTE Communications* for their support in bringing this special issue to press.

DOI: 10.3969/j. issn. 1673-5188. 2014. 03. 001
http://www.cnki.net/kcms/detail/34.1294.TN.20140822.1138.001.html, published online 22 August, 2014

Sensing, Signal Processing, and Communication for WBANs

Seyyed Hamed Fouladi[1], **Raúl Chávez-Santiago**[1,2,3], **Pål Ander Floor**[1,2,3], **Ilangko Balasingham**[1,2,3], **and Tor A. Ramstad**[1]

(1. Department of Electronics and Telecommunications, Norwegian University of Science and Technology, Trondheim, Norway;
2. The Intervention Center, Oslo University Hospital, Norway;
3. Institute of Clinical Medicine, University of Oslo, Norway)

Abstract

A wireless body area network (WBAN) enables real-time monitoring of physiological signals and helps with the early detection of life-threatening diseases. WBAN nodes can be located on, inside, or in close proximity to the body in order to detect vital signals. Measurements from sensors are processed and transmitted over wireless channels. Issues in sensing, signal processing, and communication have to be addressed before WBAN can be implemented. In this paper, we survey recent advances in research on signal processing for the sensor measurements, and we describe aspects of communication based on IEEE 802.15.6. We also discuss state-of-the-art WBAN channel modeling in all the frequencies specified by IEEE 802.15.6 as well as the need for new channel models for new different frequencies.

Keywords

wireless body area network; IEEE 802.15.6; signal processing; security; channel modeling

1 Introduction

Traditional healthcare systems can potentially be replaced by wireless body area networks (WBANs). In a WBAN, various sensors are used to sense vital signs. Patients suffering from conditions such as heart disease can be continuously monitored with a WBAN [1], [2]. Data from all the sensors is transmitted over a wireless channel to a base station, and the received measurements are processed to extract the desired information. We give an overview of sensor devices, physical layer (PHY), data link layer, and radio technology in WBAN. Different kinds of sensors can be used in a WBAN depending on requirements such as data rate and power consumption. In [3] and [4], the authors give an overview of body area networks and discuss WBAN communication types and related topics. In this work, we discuss signal processing, implant communication, and security, which have scarcely been discussed in relation to WBANs.

The IEEE 802.15.6 Task Group was established to standardize WBAN technologies and communication [5]. The main focus of the 802.15.6 standard is low-power sensors used on, in, or near the human body [5]. The standard supports physical layers, including narrowband (NB) and ultrawideband (UWB) radio interfaces, and human body communication (HBC).

In this paper, we describe wearable and implantable sensors for electrocardiography (ECG), electroencephalography (EEG), electromyography (EMG), blood pressure (BP), pulse oximetry (SpO2), accelerometer, and wireless capsule endoscopy (WCE). The measurements from these sensors need to be processed, so we discuss some of the signal-processing techniques for WBAN. We also discuss the PHYs of 802.15.6 and channel modeling in all the frequencies specified by the standard.

This paper is organized as follows: In section 2, we describe the sensors used in a WBAN and their requirements. We also discuss some important common data processing techniques for medical applications. In section 3, we discuss the PHYs in 802.15.6. In section 4, we review previous works on on-body and in-body channel modeling; in particular, we focus on implant communications. In section 5, we discuss network security. In section 6, we discuss future challenges. In section 7, we make some concluding remarks.

2 Sensors and Their Requirements

In this section, we review sensors typically used in a WBAN

The research work was performed, in part, of the research project Medical sensing, localization and communications using ultra wideband technology (MELODY) contract no. 285885, and Adaptive Security for Smart Internet of Things in eHealth (ASSET) contract no. 213131, which both are funded by the Research Council of Norway.

as well as the requirements of these sensors in terms of data rate and power consumption. We also discuss common signal-processing techniques. We consider two main categories of sensors: wearable and implantable. The first category includes ECG, EMG, EEG, accelerometer, BP, and pulse oximetry sensors. The second category includes glucose monitoring and WCE sensors.

2.1 Sensors

2.1.1 Electrocardiography

ECG is widely used in biomedical sensing, and many wireless systems for ECG monitoring have been proposed [6]. The ECG waveform shows the propagation of electric potentials through the heart muscle as a function of time. This propagation is the result of contraction of the heart muscle, and the performance of the heart can be determined by analyzing the ECG waveform. ECG measurements are based on twelve or six leads of the electrical activity of the heart. However, wireless sensors are generally only used in ambulatory scenarios to take ECG measurements, which are typically based on a subset of these leads. In [6] and [7], ECG measurements are wirelessly transmitted at a required data rate 288 kbps and 71 kbps for 12 and 6 leads, respectively.

2.1.2 Electroencephalography

The electrical activity of the brain can be monitored by EEG. Ambulatory EEG (AEEG) is valuable for diagnosing epilepsy and monitoring patient response to therapy [8]. Much of the information gained from AEEG over a 20 to 40 minute period cannot be gained from regular EEG over the same period. This has led to improved wireless EEG sensors that reduce the need for more data-intensive AEEG recording during daily activities. The required data rate for EEG is only 43.2 kbps [9].

2.1.3 Electromyography

EMG is the recording and analysis of electrical activity of skeletal muscles. The instrument used for this purpose is called an electromyograph, and the record produced is called an electromyogram [10]. EMG signals can be used to detect medical abnormalities and analyze biomechanics. EMG is often used to examine mechanisms associated with daily physical activities that induce pain and to devise related treatment regimes [11].

2.1.4 Blood Pressure

Blood pressure is one of the most important vital signs. Increased blood pressure (hypertension) increases the risk of myocardial infarction, congestive heart failure, stroke, kidney failure, and blindness. Devices that measure blood pressure are mostly based on a sphygmo-manometric obstructive arm-cuff, which is clumsy, uncomfortable, and allows only intermittent measurement every several minutes. Continuous cuffless blood pressure monitoring opens up new possibilities for hypertension diagnosis and treatment, cardiovascular event detection, and stress monitoring [12].

2.1.5 Pulse Oximetry

Pulse oximetry is a standard way of measuring arterial blood oxygen saturation (SpO2) in operating rooms, intensive care units, and pediatric care units. SpO2 is one of the most important vital signs, especially for the early detection of hypoxemia. Trauma management involves accurate monitoring of several physiological parameters, including SpO2, so that proper action can be taken to preserve critical functionality. State-of-the-art integrated circuits, wireless communications, and physiological sensing paves the way to miniature, lightweight, low power, intelligent pulse oximeters that are appropriate for WBAN applications [7].

2.1.6 Accelerometers

Accelerometers are used to measure acceleration acting on a device and convert this acceleration to an electrical signal. Algorithms can be used to classify the subject's movement into one of a few groups [13]. Research has also shown that accelerometers are effective in long-term activity monitoring and recognition. One application of accelerometers is monitoring the uncontrolled body movements (dyskinesias) of Parkinson's patients [14]. This may lead to a more effective use of levodopa, a drug used to treat the symptoms of Parkinson's disease.

2.1.7 Wireless Capsule Endoscopy

Endoscopic and radiological investigation of the small intestine had limited diagnostic operation before the year 2000. This meant that intestinal disease was sometimes diagnosed late, which worsened the patient's prognosis. WCE is a recent technique for examining the small intestine in a non-invasive way. WCE is performed using a camera the size and shape of a pill. The patient swallows the camera so that the doctor can view images of the gastrointestinal tract. However, this real-time video imaging requires a high data rate [15], [16]. It has been demonstrated WCE is more effective than other techniques in detecting small intestine diseases [17].

The patient swallows the WCE camera with water and wears a recorder belt around the waist. Some hours later, medical staff analyzes the video created from the still images transmitted from the WCE to the recorder belt. The high-quality images from the WCE camera can be analyzed in real time to enable more precise diagnosis; however, this additional capability increases transmission complexity and power consumption. The WCE camera should consume the lowest amount of power possible, on the order of microwatts, and should only be about 300 mm^3. To transmit video in real time, a high-rate communication link is needed. For VGA images, i.e., 640 × 480 p and 10 fps, 73.8 Mbps is needed [18]. Therefore, powerful, low-complexity compression algorithms are necessary to decrease data rate and power consumption.

2.2 Data Rate

Different applications in a WBAN require different data rates. For example, several megabytes per second are required for WCE whereas only several kilobytes per second is required for ECG. **Table 1** shows the required data rates for WBAN applications [7], [9], [19].

▼Table 1. Data rates for different applications

Application	Data rate	Power Consumption
ECG (12 leads)	288 kbps	Low
ECG (6 leads)	71 kbps	Low
Glucose monitoring	1600 bps	Very Low
SpO2	32 bps	Low
WCE	>2 Mbps	Low
WCE with VGA (640 × 480 p, 24 bits, 30 fps)	210.9 Mbps	Low
Blood pressure	10 bps	High
Audio	1.4 Mbps	High
EMG	320 kbps	Low
EEG	43.2 kbps	Low
Neural monitoring (512 sensors)	430 Mbps	Low

ECG: electrocardiogram
EEG: electroencephalogram
EMG: electromyography
SpO2: arterial blood-oxygen saturation
WCE: wireless capsule endoscope

2.3 Power Consumption

Power consumption is one of the most significant constraints in a WBAN. There are limitations on the size of batteries and, as a consequence, the amount of power they can provide. In some WBAN applications, such as implant sensors, the sensors need to work for several years without battery replacement. In order to save power, a thermoelectric MEMS generator can be used to scavenge energy from the surrounding environment [19]. Harvesting energy from commercial radio frequency transmissions to power WBAN nodes has been shown to be feasible [20]. Also, body heat can be used to help power wearable sensor nodes [8]. Available power in sensors is used for wireless communication, sensing, and data processing [19]. However, the main cause of power consumption in sensors is wireless communication.

A significant amount of power can be saved by modifying standard protocols according to the specific needs of WBAN application [20]. Moreover, power-aware sensor nodes can estimate the amount of transmission power needed to keep connected to the network [21]. With this concept, each node in the network can trade-off performance for energy efficiency [22].

2.4 Signal Processing

Because power is limited, low-power wireless communication and processing algorithms are needed. In a WBAN, signal processing involves processing the measurements of sensors and transmitted data to extract the desired information. Data processing within the sensor nodes must not be complex in order to reduce power consumption and prolong battery life in the sensors. However, data processing and analysis outside the sensors can be complex and power and time consuming. In this subsection, we introduce some common signal-processing algorithms for the sensors mentioned in subsection 2.1.

Sensing involves detecting a physical presence of data and transforming the signal into a format that can be read by an observer or instrument. A well-designed WBAN can give doctors accurate real-time and historical information. Data derived from different applications must be preprocessed so that it is ready for analysis. In [21], the authors propose a data preprocessing model for decreasing the power consumed during communication between nodes and for improving data transmission in wireless sensor networks. The model also provides a way of determining the integrity of data. If the data is incomplete, the error is identified, missing data is added, and corrupted data is repaired. Data processing in nodes helps eliminate redundant information, decrease the data rate, and save power. Some signal-processing algorithms are used to detect abnormalities and artifacts and track the mobile implantable sensor. Other signal-processing algorithms are used for classification.

2.4.1 Compression

Data processing usually requires less power than wireless transmission. Thus, it is important to develop algorithms that reduce the amount of information transmitted. Compression is a well-known technique used before transmission to decrease the data rate. In a wireless sensor network, such compression algorithms need to be energy-efficient. Much research has been done on data compression for wireless networks. In [23], data-compression approaches are categorized as distributed data compression and local data compression. Looking at distributed data compression, the authors of [24] aim to find a function or model that fits a best set of input measurements obtained by a specific class of sensor node. Parametric and non-parametric modeling is used. The authors of [25] use distributed transform coding to decompose data into components or coefficients, which are then coded according to their individual characteristics. There are also several well-known compression techniques, such as Karhunen–Loeve transform, cosine transformation, and wavelet transformation, which are used in image and video compression applications. Distributed source coding (DSC) is a well-known technique, based on the Slepian-Wolf theorem, for data compression in wireless sensor networks. The authors of [26] proposed energy-efficient distributed source-coding methods that have a spatial correlation for wireless sensor networks. Recently, a technique based on sampling theory was proposed by the authors of [27]. This technique, called compressed sensing, has low complexity at the sensor nodes and saves power. In other recent research, signals are reconstructed using approaches such as basis pursuit and orthogonal matching pursuit [28], [29]. In [30], an approach based on compressed sensing was proposed to compress ECG signals. Local data compression is discussed in [23] and its references.

In WCE, the diagnosis can be improved by increasing the data and frame rate. Because power consumption is limited, the data rate and frame rate should also be limited as long as the image quality satisfies hospital staff. To decrease the data rate and save power, the WCE images are compressed and encoded. A differential pulse-coded modulation (DPCM) coder requires little memory and is very easy to implement in a WBAN. The authors of [31] proposed a low-complexity algorithm for encoding WCE images with a DPCM coder and combined this with decimation, dead-zone quantization, and efficient run-length coding of the quantization indices. DPCM can also be used for other medical signals, such as ECG signals. In [32], the authors studied compression schemes for a two-node sensor network. DPCM was used to remove temporal correlation, and distributed quantization was used to exploit inter-sensor correlation. In [33], a communication system for wireless sensor networks is proposed. Low-complexity, delay direct-sum source encoder is used to reduce power consumption.

2.4.2 Localization of Capsule Endoscopy

WCE was first proposed by Given Imaging Ltd [34]. The process enables painless diagnosis within the gastrointestinal tract, specifically the small intestine. The capsule moves along the GI tract with the normal peristaltic movement of the gut and transmits the images to outside the body. Original localization approaches for WCE were based on the received signal strength (RSS) [35]. However, recent research has shown that localization accuracy can be increased in RSS-based approaches by using ultrawide bandwidth [16]. In [17], a technique based on magnetic localization was proposed for WCE. Magnetic tracking is attractive because the magnetic signal can pass through the human body without attenuation. A magnetic signal has low-static, low-frequency properties, and the capsule does not have to be in the line of sight of the magnetic sensors to be detected. In [36], the WCE localization problem is considered as a tracking problem. Maneuvers of the capsule endoscope, including any sudden stops and starts, can be tracked.

2.4.3 Detection

Detection theory is widely used in biomedical applications to find a special pattern in the signal background. Different kinds of unwanted background signals, created by the human body, deteriorate the ECG signals [34]. For example, EMG signals from muscle contraction and relaxation decreases the signal-to-noise ratio of ECG signals. Detection algorithms have been developed to eliminate noise and accurately detect the peak points in the ECG signal. In [37], the authors discuss the complexity and performance of detection schemes. In [38], the authors proposed a new pre-processing technique with Shannon energy envelope estimator to improve detection of R-peaks in ECG signals. In [39], the authors took into account both complexity and accuracy and created a detection algorithm using wavelets.

Detection algorithms are also applicable to EMG signals. The duration of EMG onset is short and must be precisely known [39]. A threshold is set by the statistical properties related to the amplitude distribution of the EMG signal. In [40], the authors use the maximum likelihood function in an algorithm for detecting the onset of muscle activity from EMG records.

3 WBAN Communications: IEEE 802.15.6

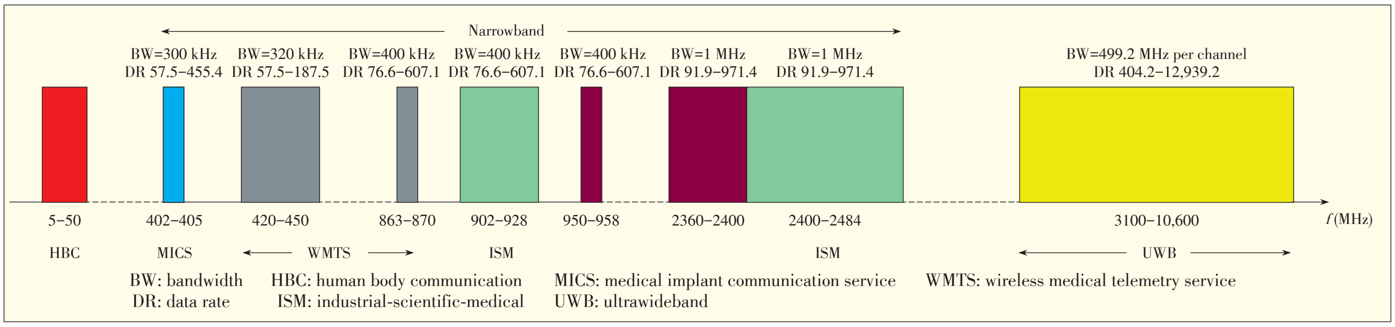
In this section, we consider the PHY layers specified in IEEE 802.15.6, including NB, UWB, and HBC. UWB and HBC PHYs are not optional whereas the NB PHY is optional [5]. Proper selection of PHYs or frequency bands is an important consideration in the development of WBANs [41]. **Fig. 1** shows the available frequency bands for WBANs [2].

3.1 NB PHY Specification

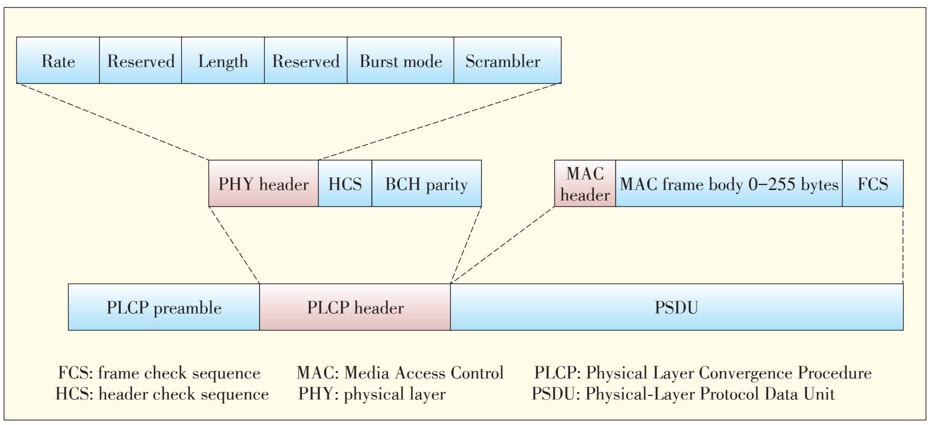
Narrowband PHY is responsible for 1) activating and deactivating the radio transceiver, 2) clear channel assessment within the current channel, and 3) data transmission and reception. NB PHY provides a way of converting a Physical Layer Service Data Unit (PSDU) into a Physical Layer Protocol Data Unit (PPDU). The PPDU frame of the NB PHY contains a physical-layer convergence procedure (PLCP) preamble, a PLCP header, and a PHY Service Data Unit (PSDU) (**Fig. 2**) [42].

The PLCP preamble helps the receiver with timing synchronization and carrier offset recovery. The other main part of the PPDU is PLCP header. The main purpose of the PLCP header is to carry essential information about the PHY parameters in order to decode the PSDU at the receiver. The PLCP header can be divided into the following fields: rate, length, burst mode, scrambler seed, reserved bits, header check sequence, and BCH parity bits. The BCH parity is responsible for improving the robustness of the PLCP header, which is transmitted after the PLCP preamble using the given header data rate in the operating frequency band. The last part of the PPDU is the PSDU. The PSDU is formed by concatenating the MAC header with the MAC frame body and frame check sequence. The PSDU is then scrambled and optionally encoded by a BCH code. The PSDU is transmitted using any of the available data rates in the operating frequency band [43].

In a WBAN, modulation has to be efficient in order to prolong life of the batteries. Gaussian frequency-shift keying, pulse-position modulation, Gaussian minimum-shift keying, differential phase-shift keying, offset quadrature phase-shift keying, and phase silence-shift keying are the predominant modulations in this area [43], [44]. These modulations reduce side lobe, are easy to implement, and are efficient in terms of



▲**Figure 1. IEEE 802.15.6 frequency bands, bandwidths, and ranges of data rates (kbps).**



▲**Figure 2. Standard PPDU structure for NB PHY.**

bandwidth.

3.2 Ultrawideband PHY Specification

UWB PHY uses both low- and high-frequency bands that are divided into 0–10 channels with a bandwidth of 499.2 MHz [5]. The low-frequency band contains the first three channels. The high-frequency band contains the remaining eight channels. Channels 1 and 6 are mandatory, and the others are optional. The central frequencies of channels 1 and 6 are 3993.6 MHz and 7978.2 MHz, respectively. However, in practice, one of these mandatory channels needs to be supported by UWB devices. The mandatory data rate is 0.4882 Mbps. The transceivers of UWB PHY are not complicated to implement and generate low-power signals. The UWB PPDU contains a synchronization header (SHR), PHY header (PHR), and PSDU. Figure 4 in [45] shows the UWB PPDU structure. The SHR includes a preamble and start frame delimiter (SFD). The PHR provides the data rate of the PSDU, length of the payload, scrambler seed, and decoding procedure in the receiver. The SHR comprises repetitions of Kasami sequences with a length of 63.

3.3 HBC PHY Specification

HBC PHY operates in two frequency bands and has a bandwidth of 4 MHz. The central frequencies of the low and high bands are 16 MHz and 27 MHz, respectively. The WBAN protocol, which specifies packet structure, modulation, preamble/SFD, etc., is identified by HBC. Figure 5 in [45] shows the PPDU structure of electrostatic field communication. This structure contains a preamble, SFD, PHR, and PSDU. The preamble and SFD are generated and sent before the packet header and payload. To ensure packet synchronization, the preamble sequence is transmitted four times. The preamble sequence is used to find the start of the packet in the receiver, and then the receiver locates the start of the frame by detecting the SFD. HBC has been referred to as intra-body communication [46]. In [46], the IBC transceiver design and mathematical models of the human body are presented.

4 Radio Propagation for Body Area Networks

Accurate channel models have to be used to fairly evaluate and compare the performance of different PHYs. Channel models have to characterize the mean path loss of WBAN devices as well as the scattering around the mean value caused by different postures of human bodies or objects in the vicinity, i.e.,

$$PL_{[dB]} = PL(d) + S \tag{1}$$

where PL is the total path loss, S is the scattering that accounts for different fading phenomena, and $PL(d)$ is the distance-dependent loss. In some WBAN scenarios, $PL(d)$ can be computed by the Friis equation given as

$$PL(d)_{[dB]} = PL_0 + 10n\log_{10}\left(\frac{d}{d_0}\right) \tag{2}$$

Where PL_0 is the path loss at a reference distance d_0 and n is the path loss exponent. Seven different propagation scenarios (S1 through S7) in which compliant WBAN devices may operate were identified in the IEEE 802.15.6 standard [47] (see **Table 2**).

These scenarios are determined according to whether the

WBAN nodes [5] are implanted inside the body; on the body surface, i.e., in contact with the skin or 2 cm from the skin; or external, i.e., between 2 cm and 5 m from the body.

Much research has been done on characterizing the WBAN channel for body-surface communication at different frequencies for NB and UWB signals. On the other hand, there are significantly fewer channel models for implant communication, in part because it is impossible to perform in-body measurements on people, and electromagnetic simulation tools are expensive.

▼Table 2. Propagation scenarios for WBAN communications [47]

Scenario	Description	Frequency Band	Channel Model
S1	Implant to implant	402–405 MHz	CM1
S2	Implant to body surface	402–405 MHz	CM2
S3	Implant to external	402–405 MHz	CM2
S4	Body surface to body surface (LOS)	13.5, 50, 400, 600, 900 MHz 2.4, 3.1–10.6 GHz	CM3
S5	Body surface to body surface (NLOS)	13.5, 50, 400, 600, 900 MHz 2.4, 3.1–10.6 GHz	CM3
S6	Body surface to external (LOS)	900 MHz 2.4, 3.1–10.6 GHz	CM3
S7	Body surface to external (NLOS)	900 MHz 2.4, 3.1–10.6 GHz	CM3

In the following, we summarize different WBAN path-loss models. We survey the literature emphasizing implant communication because this topic requires more attention.

4.1 Channel Models for Implant Communications

Scenarios S1, S2, and S3 correspond to implanted BAN nodes. Two IEEE 802.15.6 channel models, CM1 and CM2, can be used to characterize the propagation scenarios for implant nodes.

4.1.1 Narrowband Signals

CM1 and CM2 have been developed for 402–405 MHz, which is allocated to medical implant communication services in many parts of the world. This band offers good propagation through human tissue and enables the use of reasonable-size antennas [48]. However, its limited bandwidth constrains the communication devices to low transmission rates. The models are described by (1) and (2), with the scattering term being a normally distributed random variable. The parameters of these models for deeply implanted and body-surface WBAN devices can be found in [47]. An approximation of S3 can be obtained by combining S2 with S6 or S7. These models were the result of highly innovative research based on a 3D immersive visualization and simulation platform that included frequency-dependent dielectric properties of more than 300 parts of the male anatomy [49].

Other propagation models for implants using 418 MHz and 916.5 MHz were presented in [50]. In the higher frequency band, the loss was greater than expected. In [51], the propagation loss of antennas implanted in the body was computed using numerical simulations for industrial-scientific-medical (ISM) frequency bands. These bands were 433 MHz, 915 MHz, 2450 MHz, and 5800 MHz. However, the simulations were done with simplistic single-layer and triple-layer tissue structures, and no mathematical formulas for path loss were provided. In [52], numerical and experimental path loss were investigated using ingested wireless implants in 402 MHz, 868 MHz, and 2.4 GHz. Likewise, measurements were taken in phantoms (chemical solutions specially formulated to mimic the dielectric properties of human tissues), and path loss was numerically simulated for insulated dipole antennas in the ISM band at 2.457 GHz in order to derive formulas for a propagation path of up to 80 mm [53]. However, as in the previous case, the homogeneous propagation scenario was very simplistic.

More accurate channel models for implant communication in different ISM frequency bands have not yet been reported in the literature. There are opportunities to innovate in realistic anatomical voxel models and in-vivo measurements on animals to characterize path loss [54].

4.1.2 Ultrawideband Signals

There have been limited attempts to model the implant propagation channel for UWB signals. Only two models for WBAN devices implanted in the human chest have been reported. The first model [55] was developed through numerical simulations using a voxel anatomical model that included nearly 50 types of tissue with a spatial resolution of 2 mm. This model predicts a root mean square (RMS) delay spread of around 0.2 ns. The second model [56] predicts an RMS delay spread of less than 1 ns, which agrees with the results in [55], and the path loss is given as

$$PL_{[dB]} = PL_{0[dB]} + a\left(\frac{d}{d_0}\right)^n + S \qquad (3)$$

where a is a fitting constant, n is an empirical exponent, and S is a normally distributed RV. This formula does not have the same form as the Friis equation, but it is a better fit to the data obtained from numerical simulations using an anatomical model that is not homogenous and includes 24 different kinds of tissues with voxel resolution of 2 mm. A similar simulation approach as that in [56] was taken to obtain a UWB path-loss model for the abdominal region [57]. In-vivo experiments [58], [59] have demonstrated the feasibility of high-data-rate transmission over implant UWB links, and further research in this area is encouraged, especially in the characterization of frequency-dependent loss that cannot be neglected in UWB channels [60].

4.2 Channel Models for Body–Surface Communication

In 802.15.6, body surface to body surface (BS2BS) links over 5–50 MHz are established by using the human body as the communication medium. The HBC channel comprises the frequency response and noise. BS2BS in 400 MHz, 600 MHz, 900 MHz, 2.4 GHz, and 3.1–10.6 GHz are described by (2), and the corresponding parameters for hospital room and anechoic chamber measurements are found in [47]. Alternative and more detailed channel models are described in [50].

5 Security

Because many signals transported in a BAN are considered biometric data, it is important to protect this data from being observed and analyzed by unwanted parties. Encryption of sensitive data is one solution. The encryption algorithm of choice depends on how much complexity the system allows. Data from the fusion center and network coordinator can be easily encrypted using asymmetric cryptography a fusion center and network coordinator allows more complex algorithms, such as RSA, to be implemented. For sensor nodes, especially implants, the situation is different because of complexity. Symmetric cryptography solutions, especially stream ciphers, can be uncomplicated.

5.1 Cipher Security

The security of a cipher depends on several factors but is generally quantified by equivocation and work characteristic.

5.1.1 Equivocation

Equivocation is the uncertainty surrounding what was transmitted (source data) and the encryption key after an unknown party, referred to as a cryptanalyst (CA), has observed the encrypted data stream. Equivocation can be expressed in terms of conditional entropies: Let $\mathbf{X}$ denote a vector of N_x samples from the data source and $\mathbf{K}$ denote a vector of N_k key samples. Equivocation is then given by $H(\mathbf{X}|\mathbf{Y})$ and $H(\mathbf{K}|\mathbf{Y})$, where $\mathbf{Y}$ is a collection of N_x encrypted samples available to the CA. A cipher is considered secure if no information about the transmitted vector or key has been revealed when the enciphered vector has been observed, i.e., when $H(\mathbf{K}|\mathbf{Y})=H(\mathbf{K})$ and $H(\mathbf{X}|\mathbf{Y})=H(\mathbf{X})$. When these conditions are there is perfect secrecy or strongly ideal secrecy [61].

Perfect secrecy is achieved for arbitrary distribution on $\mathbf{X}$ as long as the key is uniformly distributed (completely random) over $N_k \geqslant N_x$ samples. That is, the key cannot repeat, and its length is the same as that of the data sequence transmitted. For sensors monitoring medical conditions over hours and even days, perfect secrecy is impractical.

Strongly ideal secrecy can be achieved for a key that repeats several times over the data sequence as long as that data sequence has no redundancy and is uniformly distributed.

In practice, especially with a complexity constraint, it may be difficult to remove all redundancy in the data sequence. If $N_k < N_x$, the equivocation becomes zero after the CA has gathered N_{UP} samples. This is referred to as the unicity point (UP). The UP is a function of the redundancy in a signal and increases as redundancy is reduced.

5.1.2 Work Characteristic

Although $N_x \geqslant N_{UP}$, the CA may still have problems isolating the correct solution to the cipher because this solution may be difficult to compute. The computational work that the CA must do to break the cipher is called the work characteristic. To ensure a high work characteristic, it is necessary to create mathematical problems that are known to be computationally difficult to solve, e.g., factorization of prime numbers. Applying chaotic maps prior to the encryption algorithm confuses the CA and makes it difficult to collect the statistics needed to break the cipher. Some chaotic maps, such as cat maps [62] are easy to implement.

5.2 Low-Complexity Symmetric Cryptography

A good low-complexity enciphering system should contain redundancy removal, map from given data distribution to a uniform distribution, chaotic map, and simple encryption method.

5.2.1 Simple Encryption Method

We assume that each source sample is limited so that $x_i \in [-A, A]$, and each key sample is uniformly distributed over $[-A, A]$.

Medical signals contain redundancy as correlation between consecutive samples. By removing this correlation, both compression and higher security can be achieved. There are many methods for removing correlation, some of which are mentioned in subsection 2.4.1.

A map from an arbitrary source distribution to a uniform distribution is relatively easy to create. If the source density $p_x(x)$ is monovalent and symmetric around 0, then

$$x_u = g(x) = A\int_{-x}^{x} p_x(x)\mathrm{d}x \tag{4}$$

is uniformly distributed over $[-A, A]$. For a vector, each sample can be transformed through this function. If correlation has been removed, the output will have a jointly uniform distribution at the output.

Many simple symmetric schemes can be implemented. One such scheme is the Shannon cipher [63] for encrypting analog information sources.

$$f(x+k) = \begin{cases} x+k-A & x+k \geqslant 0 \\ x+k+A & x+k < 0 \end{cases} \tag{5}$$

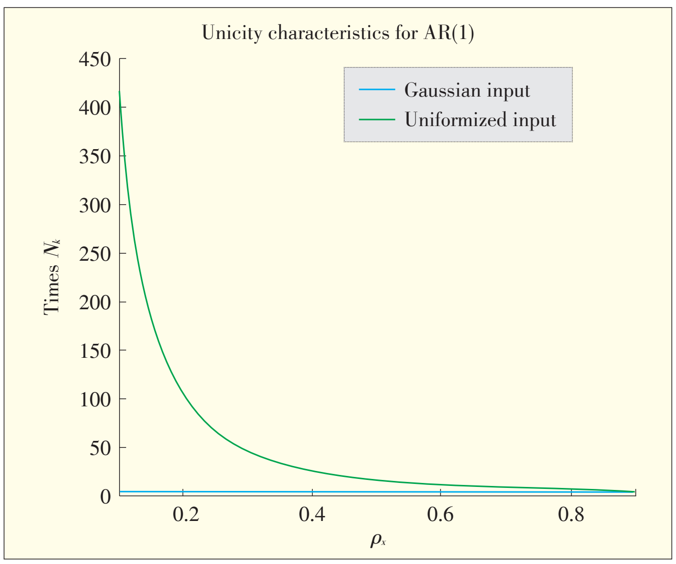
This cipher is very uncomplicated and has been shown to provide both perfect secrecy [63] and strongly ideal secrecy [64] under the conditions mentioned in subsection 5.1.1.

In a more practical situation, where there is some redundancy/structure left in the signal, the UP indicates the security of the algorithm and must be determined for relevant data. We consider a first-order Gaussian autoregressive process AR(1) with correlation ρ_x between consecutive samples. If each sample is mapped through (4), the lower bound of the UP is

$$N_{UP} \geqslant \frac{2H(\mathbf{K})}{-\ln\left(1-\rho_x^2\right)}+1 \tag{6}$$

Fig. 3 shows the lower bound of the UP for AR(1) as a function of correlation when 1) the signal (blue) is encrypted directly, and 2) the transformed vector (green) is encrypted.

By reducing correlation down to about 0.2, which is possible in practice, the key has to repeat at least 100 times before the CA can break the cipher. Direct encryption is insufficient: the key can only repeat a couple of times before UP is reached independent of ρ_x. It is important to map the data sequence through (4) before encryption.



▲**Figure 3. Key repetition rate as a function of correlation.**

5.2.2 Security and Fidelity

If Shannon's cipher (5) is applied to noisy channels, both source and key must be quantized prior to encryption; otherwise, large decoding errors will corrupt the reconstructed signal. The larger the noise is, the larger the quantizer step-size should be. The fidelity in the reconstruction is therefore reduced whenever the noise increases.

In [65], the authors argue that there is a tradeoff between security and fidelity when analog signals are encrypted. Fidelity drops because of increased quantizer step size, and this means that more secure encryption is possible because fewer samples need to be transmitted when the signal is heavily compressed. Another reason why there is a tradeoff between security and fidelity when $N_k < N_x$ is that the transform in (4), which is optimal from a security perspective, is not necessarily the optimal transform with respect to quantization and transmission.

5.2.3 Key Distribution

A problem with symmetric key cryptography is secure distribution of the key(s) being used. However, for medical BAN configured at the hospital, several keys can be pre-stored and changed by a simple rule decided by the person responsible for installing the medical sensors. If the UP is high, relatively few keys need to be installed. If a high enough UP is not possible, chaotic maps can be implemented to increase the work characteristic.

6 Future Challenges

Recent advancements in microelectronics make it possible to have a WBAN with numerous sensor nodes, and high-speed connectivity is possible with multiple antennas. As WBANs become bigger, they will encounter problems in terms of multi-hop routing, end-to-end delay, and service provisioning with priority- transmission. All these features will have to be carefully optimized in a system with constraints such as small footprint, limited power, and limited computational possibilities.

WBANs can be arbitrarily deployed, and this leads to interference. For example, passengers with WBANs in a confined area, such as a bus, will experience mutual interference. It has to be determined whether cognitive radio networking technology can help mitigate interference in WBANs [62]. Another interesting research problem is the possibility of using WBAN-to-WBAN interaction to improve connectivity and communication service. However, for this to be achieved, it has to be determined whether different WBANs can communicate and exchange private (possibly sensitive) medical information for the mutual benefit of the WBANs. Privacy and access control for WBAN-to-WBAN communications both have to be extensively researched. Multiple sensors operating at different rates and having different priorities in terms of mobility raise difficult problems. Such a system needs to treat PHY, MAC, and network layers in a cross-layer manner.

7 Conclusion

In this paper, we have described the different aspects of a WBAN, including sensors and their requirements in terms of data rate and power consumption. We have discussed signal-processing techniques, such as compression, detection and localization, in mobile implant sensors. The PHY layers of 802.15.6, and body-surface and in-body channel modeling were described for the frequencies specified in 802.15.6. Possible security techniques for WBANs were also discussed.

References

[1] S. Ullah, H. Higgins, B. Braem, B. Latre, C. Blondia, I. Moerman, *et al.*, "A comprehensive survey of wireless body area networks," *J. of Medical Syst.*, vol. 36, no. 3, pp. 1065–1094, Jun. 2012. doi: 10.1007/s10916-010-9571-3.
[2] B. Latré, B. Braem, I. Moerman, C. Blondia, and P. Demeester, "A survey on wireless body area networks," *Wireless Networks*, vol. 17, no. 1, pp. 1–18, Jan. 2011. doi: 10.1007/s11276-010-0252-4.
[3] M. Chen, S. Gonzalez, A. Vasilakos, H. Cao, and V. C. Leung, "Body area networks: A survey," *Mobile Networks and Applicat.*, vol. 16, no. 2, pp. 171–193,

Apr. 2011. doi: 10.1007/s11036-010-0260-8.

[4] R. Cavallari, F. Martelli, R. Rosini, C. Buratti, and R. Verdone, "A survey on wireless body area networks: technologies and design challenges," *IEEE Commun. Surveys & Tutorials*, to be published.

[5] I. S. Association, "IEEE standard for local and metropolitan area networks-part 15.6: wireless body area networks," *IEEE Standard for Inform. Technology*, vol. 802, pp. 1–271, 2012.

[6] M. M. Baig, H. Gholamhosseini, and M. J. Connolly, "A comprehensive survey of wearable and wireless ECG monitoring systems for older adults," *Medical & biological Eng. & computing*, vol. 51, no. 5, pp. 485–495, May 2013. doi: 10.1007/s11517-012-1021-6.

[7] S. Arnon, D. Bhastekar, D. Kedar, and A. Tauber, "A comparative study of wireless communication network configurations for medical applications," *IEEE Wireless Commun.*, vol. 10, no. 1, pp. 56–61, Feb. 2003. doi: 10.1109/MWC.2003.1182112.

[8] B. Gyselinckx, C. Van Hoof, J. Ryckaert, R. F. Yazicioglu, P. Fiorini, and V. Leonov, "Human++: autonomous wireless sensors for body area networks," in *Proc. of the IEEE 2005 on Custom Integrated Circuits Conf.*, 2005 San Jose, USA, pp. 13–19. doi: 10.1109/CICC.2005.1568597.

[9] J. F. Rizzo, J. L. Wyatt Jr, and L. Theogarajan, "Minimally invasive retinal prosthesis," US 6976998 B2, Dec. 20, 2005.

[10] J. Viby-Mogensen, "Neuromuscular monitoring," *Current Opinion in Anesthesiology*, vol. 14, no. 6, pp. 655–659, Dec. 2001.

[11] P. Bonato, P. J. Mork, D. M. Sherrill, and R. H. Westgaard, "Data mining of motor patterns recorded with wearable technology," *IEEE Eng. Med. Biol. Mag.*, vol. 22, no. 3, pp. 110–119, 2003. doi: 10.1109/MEMB.2003.1213634.

[12] F. Adochiei, C. Rotariu, R. Ciobotariu, and H. Costin, "A wireless low-power pulse oximetry system for patient telemonitoring," in *7th Int. Symp. on Advanced Topics in Elect. Eng. (ATEE)*, Bucharest, Romania, May 2011, pp. 1–4.

[13] T. R. Burchfield and S. Venkatesan, "Accelerometer-based human abnormal movement detection in wireless sensor networks," in *HealthNet'07*, San Juan, Puerto Rico, pp. 67–69. doi: 10.1145/1248054.1248073.

[14] N. L. Keijsers, M. W. Horstink, and S. C. Gielen, "Automatic assessment of levodopa-induced dyskinesias in daily life by neural networks," *Movement disorders*, vol. 18, no. 1, pp. 70–80, 2003.

[15] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*, New York: John Wiley & Sons, 2004.

[16] B. Moussakhani, J. T. Flåm, S. Støa, I. Balasingham, and T. Ramstad, "On localisation accuracy inside the human abdomen region," *IET Wireless Sensor Syst.*, vol. 2, no. 1, pp. 9–15, 2012.

[17] C. Hu, M. Li, S. Song, R. Zhang, and M.-H. Meng, "A cubic 3-axis magnetic sensor array for wirelessly tracking magnet position and orientation," *IEEE Sensors J.*, vol. 10, no. 5, pp. 903–913, May 2010. doi: 10.1109/JSEN.2009.2035711.

[18] R. Chávez-Santiago, A. Khaleghi, I. Balasingham, and T. A. Ramstad, "Architecture of an ultra wideband wireless body area network for medical applications," in *2nd Int. Symp. Applied Sciences in Biomedical and Commun. Technologies*, Bratislava, Slovakia, 2009, pp. 1–6. doi: 10.1109/ISABEL.2009.5373624.

[19] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–114, Aug. 2002. doi: 10.1109/MCOM.2002.1024422.

[20] V. Leonov, P. Fiorini, S. Sedky, T. Torfs, and C. Van Hoof, "Thermoelectric MEMS generators as a power supply for a body area network," in *13th Int. Conf. Solid-State Sensors, Actuators and Microsystems, 2005. Digest of Tech. Papers*, Seoul, South Korea, vol. 1, pp. 291–294. doi: 10.1109/SENSOR.2005.1496414.

[21] M. Wang, J. Cao, M. Liu, B. Chen, Y. Xu, and J. Li, "Design and implementation of distributed algorithms for WSN-based structural health monitoring," *Int. J. Sensor Networks*, vol. 5, no. 1, pp. 11–21, Feb. 2009. doi: 10.1504/IJSNET.2009.023312.

[22] S. Xiao, A. Dhamdhere, V. Sivaraman, and A. Burdett, "Transmission power control in body area sensor networks for healthcare monitoring," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 1, pp. 37–48, Jan. 2009. doi: 10.1109/JSAC.2009.090105.

[23] T. Srisooksai, K. Keamarungsi, P. Lamsrichan, and K. Araki, "Practical data compression in wireless sensor networks: A survey," *J. Network and Computer Applicat.*, vol. 35, no. 1, pp. 37–59, Jan. 2012. doi: 10.1016/j.jnca.2011.03.001.

[24] A. Oka and L. Lampe, "Energy efficient distributed filtering with wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 2062–2075, May 2008. doi: 10.1109/TSP.2007.911496.

[25] J. A. Gubner, "Distributed estimation and quantization," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, Jul. 1993. doi: 10.1109/18.243470.

[26] J. Chou, D. Petrovic, and K. Ramachandran, "A distributed and adaptive signal processing approach to reducing energy consumption in sensor networks," in *INFOCOM 2003*, San Francisco, USA, vol. 2, pp. 1054–1062. doi: 10.1109/INFCOM.2003.1208942.

[27] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006. doi: 10.1109/TIT.2006.871582.

[28] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2210–2219, May 2008. doi: 10.1109/TIT.2008.920190.

[29] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007. doi: 10.1109/TIT.2007.909108.

[30] L. F. Polania, R. E. Carrillo, M. Blanco-Velasco, and K. E. Barner, "Compressed sensing based method for ECG compression," in *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, 2011, pp. 761–764. doi: 10.1109/ICASSP.2011.5946515.

[31] A. N. Kim, T. A. Ramstad, and I. Balasingham, "Very low complexity low rate image coding for the wireless endoscope," in *Proc. 4th Int. Symp. Applied Sciences in Biomedical and Commun. Technologies*, 2011, p. 90. doi: 10.1145/2093698.2093788.

[32] P. A. Floor, I. Balasingham, T. A. Ramstad, E. Meurville, and M. Peisino, "Compression schemes for in-body and on-body UWB sensor networks," in *3rd Int. Symp. Applied Sciences in Biomedical and Commun. Technologies (ISABEL)*, Rome, Italy, 2010, pp. 1–5. doi: 10.1109/ISABEL.2010.5702843.

[33] H. T. Nguyen, T. A. Ramstad, and I. Balasingham, "Wireless sensor communication system based on direct-sum source coder," *IET Wireless Sensor Syst.*, vol. 1, pp. 96–104, 2011. doi: 10.1049/iet-wss.2010.0094.

[34] L. Smital, M. Vítek, J. Kozumplík, and I. Provaznik, "Adaptive Wavelet Wiener Filtering of ECG Signals," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 2, pp. 437–445, Feb. 2013. doi: 10.1109/TBME.2012.2228482.

[35] K. A. Germansky and D. A. Leffler, "Best practice & research clinical gastroenterology," *Best Practice & Research Clinical Gastroenterology*, vol. 25, pp. 387–395, 2011.

[36] B. Moussakhani, R. Chavez-Santiago, and I. Balasingham, "Multi model tracking for localization in wireless capsule endoscopes," in *Proc. 4th Int.l Symp. Applied Sciences in Biomedical and Commun. Technologies*, Spain, 2011, p. 159. doi: 10.1145/2093698.2093857.

[37] Y.-J. Min, H.-K. Kim, Y.-R. Kang, G.-S. Kim, J. Park, and S.-W. Kim, "Design of Wavelet-Based ECG Detector for Implantable Cardiac Pacemakers," *IEEE Trans. Biomed. Circuits Syst.*, vol. 7, no. 4, pp. 426–436, Aug. 2013. doi: 10.1109/TBCAS.2012.2229463.

[38] M. S. Manikandan and K. P. Soman, "A novel method for detecting R-peaks in electrocardiogram (ECG) signal," *Biomedical Signal Processing and Control*, vol. 7, no. 2, pp. 118–128, Mar. 2012. doi: 10.1016/j.bspc.2011.03.004.

[39] S. Kadambe, R. Murray, and G. F. Boudreaux-Bartels, "Wavelet transform-based QRS complex detector," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 7, pp. 838–848, Jul. 1999. doi: 10.1109/10.771194.

[40] A. P. Stylianou, C. W. Luchies, and M. F. Insana, "EMG onset detection using the maximum likelihood method," in *Proc. 2003 Summer Bioengineering Conf.*, Key Biscayne, USA, pp. 1075–1076.

[41] M. Hernandez and R. Miura, "Coexistence of IEEE Std 802.15. 6 TM-2012 UWB-PHY with other UWB systems," in *IEEE Int. Conf. Ultra-Wideband (ICUWB)*, Syracuse, USA, 2012, pp. 46–50. doi: 10.1109/ICUWB.2012.6340496.

[42] S. Ullah, M. Mohaisen, and M. A. Alnuem, "A review of ieee 802.15. 6 MAC, PHY, and security specifications," *Int. J. Distributed Sensor Networks*, vol. 2013, 2013. doi: dx.doi.org/10.1155/2013/950704.

[43] B. Choi, B. Kim, S. Lee, K. Wang, Y. Kim, and D. Chung, "Narrowband physical layer design for WBAN system," in *First Int. Conf. Pervasive Computing Signal Processing and Applicat. (PCSPA)*, Harbin, China, 2010, pp. 154–157. doi: 10.1109/PCSPA.2010.46.

[44] H. T. Nguyen, T. A. Ramstad, and I. Balasingham, "Coded pulse position modulation communication system over the human abdominal channel for medical wireless body area networks," in *IEEE 23rd Int. Symp. Personal Indoor and Mobile Radio Commun. (PIMRC)*, Sydney, Australia, 2012, pp. 1992–1996. doi: 10.1109/PIMRC.2012.6362680.

[45] K. S. Kwak, S. Ullah, and N. Ullah, "An overview of IEEE 802.15. 6 standard," in *3rd Int. Symp. Applied Sciences in Biomedical and Communication Technologies (ISABEL)*, Rome, Italy, 2010, pp. 1–6. doi: 10.1109/ISA-

BEL.2010.5702867.
[46] M. Seyedi, B. Kibret, D. T. Lai, and M. Faulkner, "A survey on intrabody communications for body area network applications," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 3, pp. 2067-2079, Aug. 2013. doi: 10.1109/TBME.2013.2254714.
[47] K. Y. Yazdandoost and K. Sayrafian-Pour, "Channel model for body area network (BAN)," *IEEE P802*, vol. 15, 2009.
[48] P. Soontornpipit, C. M. Furse, and Y. C. Chung, "Design of implantable microstrip antenna for communication with medical implants," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 8, pp. 1944-1951, Aug. 2004.
[49] K. Sayrafian-Pour, W.-B. Yang, J. Hagedorn, J. Terrill, K. Y. Yazdandoost, and K. Hamaguchi, "Channel models for medical implant communication," *Int. J. Wireless Inform. Networks*, vol. 17, no. 3-4, pp. 105-112, Dec. 2010. doi: 10.1007/s10776-010-0124-y.
[50] W. G. Scanlon, B. Burns, and N. E. Evans, "Radiowave propagation from a tissue-implanted source at 418 MHz and 916.5 MHz," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 4, pp. 527-534, Apr. 2000. doi: 10.1109/10.828152.
[51] J. Gemio, J. Parron, and J. Soler, "Human body effects on implantable antennas for ISM bands applications: models comparison and propagation losses study," *Progress in Electromagnetics Research*, vol. 110, pp. 437-452, 2010. doi: 10.2528/PIER10102604.
[52] A. Alomainy and Y. Hao, "Modeling and characterization of biotelemetric radio channel from ingested implants considering organ contents," *IEEE Trans. Antennas Propag.*, vol. 57, no. 4, pp. 999-1005, Apr. 2009. doi: 10.1109/TAP.2009.2014531.
[53] D. Kurup, W. Joseph, G. Vermeeren, and L. Martens, "Path loss model for in-body communication in homogeneous human muscle tissue," *Electronics letters*, vol. 45, no. 9, pp. 453-454, 2009.
[54] R. Chavez-Santiago, K. Sayrafian-Pour, A. Khaleghi, K. Takizawa, J. Wang, I. Balasingham, and H-B Li, "Propagation models for IEEE 802.15. 6 standardization of implant communication in body area networks," *IEEE Commun. Mag.*, vol. 51, no. 8, Aug. 2013. doi: 10.1109/MCOM.2013.6576343.
[55] J. Wang and Q. Wang, "Channel modeling and BER performance of an implant UWB body area link," in *2nd Int. Symp. Applied Sciences in Biomedical and Commun. Technologies*, Bratislava, Slovakia, 2009, pp. 1-4. doi: 10.1109/ISABEL.2009.5373707.
[56] A. Khaleghi, R. Chávez-Santiago, and I. Balasingham, "Ultra-wideband statistical propagation channel model for implant sensors in the human chest," *IET Microwaves, Antennas & Propagation*, vol. 5, no. 15, pp. 1805-1812, Dec. 2011. doi: 10.1049/iet-map.2010.0537.
[57] S. Støa, R. Chavez-Santiago, and I. Balasingham, "An ultra wideband communication channel model for the human abdominal region," in *IEEE GLOBECOM Workshops (GC Wkshps)*, Miami, USA, 2010, pp. 246-250. doi: 10.1109/GLOCOMW.2010.5700319.
[58] R. Chavez-Santiago, I. Balasingham, J. Bergsland, W. Zahid, K. Takizawa, R. Miura, and H-B Li, "Experimental implant communication of high data rate video using an ultra wideband radio link," in *35th Annu. Int. Conf. IEEE on Eng. in Medicine and Biology Society (EMBC)*, Osaka, Japan, 2013, pp. 5175-5178. doi: 10.1109/EMBC.2013.6610714.
[59] D. Anzai, K. Katsu, R. Chavez-Santiago, Q. Wang, D. Plettemeier, J. Wang, and I. Balasingham, "Experimental evaluation of implant UWB-IR transmission with living animal for body area networks," *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 1, pp. 183-192, Jan. 2014. doi: 10.1109/TMTT.2013.2291542.
[60] A. Khaleghi, R. Chávez-Santiago, and I. Balasingham, "An improved ultra wideband channel model including the frequency-dependent attenuation for in-body communications," in *Annu. Int. Conf. IEEE on Eng. in Medicine and Biology Society (EMBC)*, San Diego, USA, 2012, pp. 1631-1634. doi: 10.1109/EMBC.2012.6346258.
[61] C. E. Shannon, "Communication Theory of Secrecy Systems," *Bell Syst. Tech. J.*, vol. 28, no. 4, pp. 656-715, 1949.
[62] R. Chávez-Santiago, K. E. Nolan, O. Holland, L. De Nardis, J. M. Ferro, N. Barroca, L. M. Borges, F. J. Velez, V. Goncalves, and I. Balasingham, "Cognitive radio for medical body area networks using ultra wideband," *IEEE Wireless Commun.*, vol. 19, no. 4, pp. 74-81, Aug. 2012. doi: 10.1109/MWC.2012.6272426.
[63] C. E. Shannon, "Analogue of the Vernam system for continuous time series," in *Claude Elwood Shannon: Collected Papers, Memorandum MM*, Los Alamitos, USA: IEEE Computer Society Press, 1943, pp. 43-110.
[64] P. A. Floor, I. Balasingham, and T. A. Ramstad, "Analysis of the Shannon-Gersho (SG) cipher," internal document.
[65] A. Gersho, "Perfect secrecy encryption of analog signals," *IEEE J. Sel. Areas Commun.*, vol. 2, no. 3, pp. 460-466, May 1984. doi: 10.1109/JSAC.1984.1146071.

Manuscript received: 2014-04-04

Biographies

Seyyed Hamed Fouladi (hamed.fouladi@iet.ntnu.no) received his B.S. degree in electrical engineering at the Department of Electrical Engineering, Shahed University, Tehran, Iran in 2009 and M.Sc degree in Communication Systems at the Department of Electrical Engineering, Amirkabir University of Technology, Tehran, Iran, in 2012. He is currently a PhD student at Norwegian University of Science and Technology (NTNU), Trondheim, Norway. His research interests include statistical signal and image processing, multi-resolution signal analysis, blind signal processing, statistical modeling, detection and estimation.

Raúl Chávez-Santiago (raul.chavez-santiago@rr-research.no) graduated as an Electronics and Telecommunications Engineer at the National Polytechnic Institute, Mexico, in 1997. In 2001 he obtained a M.Sc. degree in Electrical Engineering at CINVESTAV, Mexico. He received the Ph.D. degree in Electrical and Computer Engineering from Ben-Gurion University of the Negev, Israel, in 2007. Thereafter he held a postdoctoral position at the University Paris-Sud XI, France, where he investigated radio resource management for OFDMA systems. He later held a second postdoctoral position at Bar-Ilan University, Israel. There, he researched the information theory aspects of ad hoc and cognitive radio networking. He joined the Intervention Centre, Oslo University Hospital, Norway, in 2009, where he currently investigates short-range radio communication technologies for body area network (BAN) solutions. His research is focused on implant communications and ultra wideband (UWB) technology. He is a Management Committee member in various European COST Actions.

Pål Anders Floor (andflo@rr-research.no) received his B.Sc. degree from Gjøvik University College (HIG), Norway in 2001, his M.Sc degree in 2003 and his PhD, degree in 2008, both from the Department of Electronics and Telecommunications, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. All three degrees are in electrical engineering. He was working as a Post. Doc. at the Intervention Centre at Oslo University Hospital and at the Institute of Clinical Medicine at University of Oslo from 2008 to 2013. He is currently a Post. Doc at NTNU. His research interests include joint source-channel coding, information theory and signal processing applied on point-to-point links, in small and large networks, as well as in Neuroscience, and lightweight cryptography solutions for medical BAN.

Ilangko Balasingham (ilangkob@medisin.uio.no) received the M.Sc. and Ph.D. degrees from the Department of Electronics and Telecommunications, Norwegian University of Science and Technology (NTNU), Trondheim, Norway in 1993 and 1998, respectively, both in signal processing. He performed his Master's degree thesis at the Department of Electrical and Computer Engineering, University of California Santa Barbara, USA. From 1998 to 2002, he worked as a Research Scientist at Fast Search & Transfer ASA, Oslo, Norway, which is now part of Microsoft Inc. Since 2002 he has been with the Intervention Center, Oslo University Hospital, Oslo, Norway, where he heads the Wireless Sensor Network Research Group. He was appointed as a Professor in Signal Processing in Medical Applications at NTNU in 2006. His research interests include wireless body sensor networks, microwave sensing and imaging, short range localization and tracking, and nano-neural communication networks. He has authored or co-authored 168 papers and has been active in organizing special sessions, workshops, and conferences.

Tor A. Ramstad (ramstad@iet.ntnu.no) is Professor Emeritus at the Norwegian University of Science and Technology (NTNU formerly NTH), Norway. He got his MSc PhD degrees from NTH in 1968 and 1972, where he was assistant and associate professor until he became a full professor of Communication Theory in 1983, and retired in 2012. He has been a visiting professor at UCSB, Georgia Tech, and Eurecom, France. He was Associate Editor of IEEE Acoustics, Speech and Signal Processing, and chair of the IEEE Signal Processing Workshop in 1996. He is a member of the Norwegian Academy of Technological Sciences, and was awarded the Honorary Price from the Norwegian Signal Processing Society. Professor Ramstad has had a leading role in establishing and developing the field of digital signal processing in Norway. His main interests include efficient digital filtering methods and implementations, signal compression of speech, images and video, signal interpolation for sample rate conversion, and joint source-channel coding.

DOI: 10.3969/j. issn. 1673-5188. 2014. 03. 002
http://www.cnki.net/kcms/detail/34.1294.TN.20140821.0900.001.html, published online 21 August, 2014

MAC Layer Resource Allocation for Wireless Body Area Networks

Qinghua Shen[1], Xuemin (Sherman) Shen[1], Tom H. Luan[2], and Jing Liu[3]
(1. Department of Electrical and Computer Engineering, University of Waterloo, ON N2L 3G1, Canada;
2. School of Information Technology, Deakin University, Melbourne, Australia;
3. Department of Electronic Engineering, Shanghai Jiao Tong University, China)

Abstract

Wireless body area networks (WBANs) can provide low-cost, timely healthcare services and are expected to be widely used for e-healthcare in hospitals. In a hospital, space is often limited and multiple WBANs have to coexist in an area and share the same channel in order to provide healthcare services to different patients. This causes severe interference between WBANs that could significantly reduce the network throughput and increase the amount of power consumed by sensors placed on the body. Therefore, an efficient channel-resource allocation scheme in the medium access control (MAC) layer is crucial. In this paper, we develop a centralized MAC layer resource allocation scheme for a WBAN. We focus on mitigating the interference between WBANs and reducing the power consumed by sensors. Channel and buffer state are reported by smartphones deployed in each WBAN, and channel access allocation is performed by a central controller to maximize network throughput. Sensors have strict limitations in terms of energy consumption and computing capability and cannot provide all the necessary information for channel allocation in a timely manner. This deteriorates network performance. We exploit the temporal correlation of the body area channel in order to minimize the number of channel state reports necessary. We view the network design as a partly observable optimization problem and develop a myopic policy, which we then simulate in Matlab.

Keywords

medium access control (MAC); wireless body area networks (WBANs); resource allocation; interference mitigation; partially observable optimization

1 Introduction

It is widely recognized that current hospital-centric healthcare services do not efficiently meet the needs of an ageing population. A promising solution is e-healthcare, which involves using ICT to support healthcare practices [1]. The key part of an e-healthcare system is a wireless body-area network (WBAN) comprising multiple sensors that monitor the physical condition of patients [2]. A WBAN enables continuous, remote monitoring of patients, and this increases the efficiency of medical staff and reduces the cost of hospital healthcare [3]. With an e-healthcare system, medical staff can limit the number of unnecessary visits to a patient. Data collected by the sensors enables medical staff to keep track of the condition of individual patients and react to emergencies in advance. For example, signs of cardiac arrest could be detected hours in advance, and a life-threatening situation averted [4].

However, there are still fundamental challenges to the widespread use of sensor-based WBAN in hospitals and the provision of guaranteed communication services for critical medical traffic. If medical staff are to respond rapidly, patient vital signs such as heart rate, blood pressure, respiratory rate, temperature, pulse oximetry, and level of consciousness all need to be monitored in real time. This means that the data must be transmitted accurately and without unacceptable delay [5], [6]. In addition, sensors on the body typically have limited power. This means that underlying communication protocols must be efficient, i.e., transmission errors and retransmission must be kept to a minimum. Because sensors have limited computing capability and data buffer, real-time data that cannot be transmitted within a given period is dropped, and this leads to a high report-dropping ratio. An e-healthcare network demands more efficient communication because of the critical nature of medical traffic and the limited resources of sensors. Also, because floor space is usually quite limited in a hospital, multiple WBANs for different groups of patients are often located in

This research has been supported by a research grant from the Natural Science and Engineering Research Council (NSERC) under grant No. CRDPJ 419147-11, and Care In Motion Inc., Canada.

the same area, and interference between WBANs can be severe. Therefore, MAC layer resource allocation has to be efficient and provide the high QoS needed for e-healthcare systems.

In this paper, we exploit the temporal channel correlations around human bodies to develop a centralized MAC protocol for e-healthcare systems. Specifically, we consider a scenario where multiple WBANs exist in an area, such as a ward, and contend the channel for transmission. Each WBAN includes sensors on the body that continuously transmit collected data to the Internet (via a data sink) and to the smartphone of the patient. Because of variations in the body-area channel, the transmission link between a sensor and smartphone may not be always available. To maximize network throughput and reduce the packet-drop ratio, it might seem reasonable to permit the WBAN with good channel quality and cached data in sensors to transmit. However, this means that the real-time channel and buffer state information of all WBANs is required. Because sensors on the body have limited computing and energy resources, this information cannot always be accurately measured and provided by the sensors. This leads to inefficient use of channels.

To address this issue, we use the temporal channel correlations to guide MAC resource allocation. Specifically, we represent the channel state with a belief state and use this metric to allocate access to the channel. The belief state, which is not chosen for transmission, is updated according to the statistical information. Only one sensor from a WBAN needs to report its current state so that channel-state reports are minimized. Given the incomplete nature of network state information, we treat the throughput-maximization problem as a partly observable optimization problem. We first analyze the dynamics of the belief states and buffer states. Then, we create a myopic policy and investigate its drawback in terms of incurred packet dropping. Then, we propose a modified myopic policy in which the future impact of a current decision is approximated. Finally, we compare our proposed policy with Round Robin (RR) to demonstrate its effectiveness.

The remainder of this paper is organized as follows. In section 2, we discuss related works. In section 3, we discuss the system model to be studied. In section 4, we formulate the problem. In section 5, we discuss the policy design. In section 6, we give simulation results. In section 7, we conclude and discuss future research directions.

2 Related Works

In this section, we review works on resource allocation in the MAC layer of a WBAN as well as works that cover the problem of partly observable optimal control.

Resource allocation in the WBAN MAC layer has long been an important research topic. In [7], a fuzzy logic algorithm is used in a hospital environment to adjust the MAC layer control parameters according to real-time network information. To support the transmission of medical traffic within coexisting WBANs, IEEE 802.15.6 has been proposed. The standard describes collaborative and non-collaborative methods, such as beacon shifting and channel hopping, for eliminating interference between WBANs [8], [9]. However, it does not specify how to use these methods with different network settings. Interference-mitigation schemes for WLANs, including the busy tone scheme [10]–[13], are not suitable for WBANs because energy-consuming control signals quickly drain the power of sensors. To reduce the amount of energy consumed by sensors, the scheduling problem can be formulated using game theory, and heuristic cooperation can be used in the scheduling policy [14]. In [15], a network that can tolerate concurrent transmission of multiple WBANs is described. The authors propose an uncomplicated scheduling scheme inspired by the random incomplete coloring scheme. However, in all of these mentioned works, variations in the body area channel are not considered.

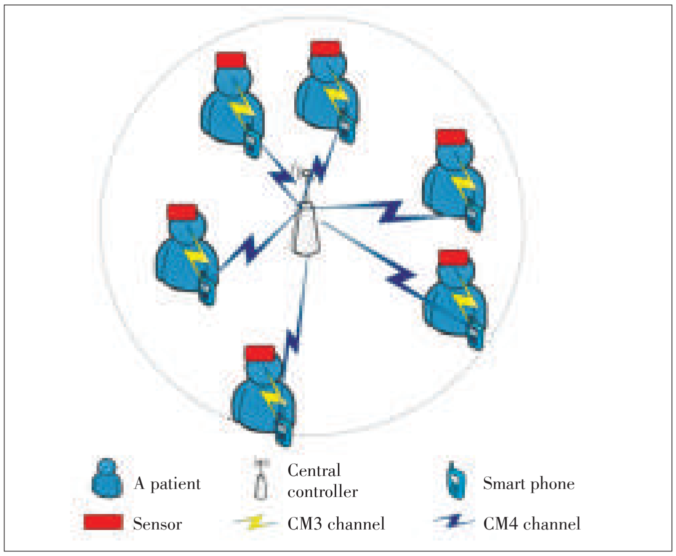
In this paper, we describe the throughput maximization problem as a partly observable optimization problem because of the incomplete nature of network state information. The partly observable optimization problem has been studied extensively. Research on this topic started with scheduling over a single random process. In [16] and [17], a random process governed by a Markov chain is considered. The authors show that the conditional probability distribution of the current state (given previous control decisions and observations) is sufficient statistical information for an optimal scheduling policy. Moreover, the convex property of a corresponding value function is proved [16]. This is the key to obtaining an optimal policy. In [18], the authors studied optimal policies for scheduling over multiple random processes.

3 System Model

In this section, we describe the system model in terms of network, channel, and traffic and then describe the channel-access scheme.

3.1 Network Model

We consider a network comprising N_w WBANs and one central controller (**Fig. 1**). The central controller allocates channel resources to WBANs and forwards the medical information through a wired network. Each WBAN corresponds to a single patient. Within each WBAN, there is a smart phone and one sensor. In this paper, only one sensor is considered because a single sensor can monitor most vital signals nowadays [19]. The smart phone collects information from the sensor in the same WBAN and transmits this information to the central controller. A sensor on the body has limited power and computing capacity whereas a smartphone has greater power supply and computing capacity. Sensors are set to turn a radio on at a predefined time but are in sleep mode most of time. In contrast,



▲Figure 1. System model.

the smartphone is always turned on. The timing of the system is partitioned into slots, and the duration of each of these slots is denoted T.

3.2 Channel Model

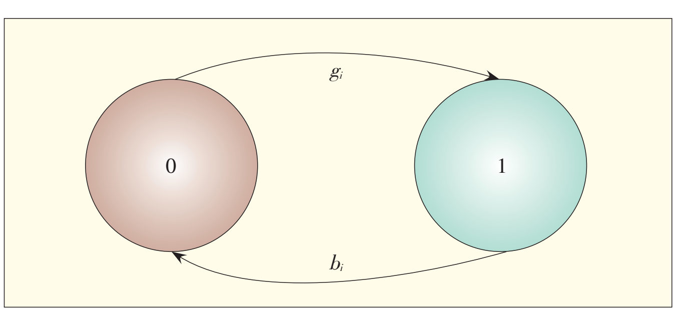
Medical traffic is transmitted in two hops from the sensor to the central controller (Fig. 1). As in IEEE 802.15.6 [20], [21], we denote the first-hop body-surface-to-body-surface channel CM3 and the second-hop body-surface-to-external channel CM4. The CM4 channel is modeled as free space wireless channel, and the transmissions of the second hop are error-free because of the ample transmission power and clear channel conditions. CM3 has severe and varying path loss due to the absorption of human body. There are also strong temporal correlations of channel between neighboring time slots.

To obtain the features of CM3 without any loss of generality, we use the Gilbert Elliot (GE) model (**Fig. 2**) [22], [23]. In this model there are two channel states: on (error-free transmission) and off (unsuccessful transmission).

Let $C_i(n)$ denote CM3 channel state for the ith WBAN over the nth time slot. If the channel is on, $C_i(n) = 1$; otherwise $C_i(n) = 0$. Let R_i^c and Π_c^i denote the probability transition matrix and stationary distribution of CM3 for the ith WBAN, respectively. According to the GE model, R_i^c can be given as

$$R_i^c = \begin{bmatrix} 1-g_i & g_i \\ b_i & 1-b_i \end{bmatrix} \quad (1)$$

where g_i is the conditional probability of the channel changing from off to on, and $g_i \triangleq \Pr\{C_i(n)=1|C_i(n-1)=0\}$ for $n \in \{1, 2,...\}$. The conditional probability of the channel changing from on to off is given by b_i, and $b_i \triangleq \Pr\{C_i(n)=0|C_i(n-1)=1\}$ for $n \in \{1, 2,...\}$. The corresponding stationary distribution is given by $\Pi_c^i = [b_i/(b_i+g_i), g_i/(b_i+g_i)]$. In the GE model, if a channel tends to stay in its current state, the channel is positively correlated, i.e., $1 > b_i + g_i$.



▲Figure 2. On/off wireless channel model (Gilbert Elliot model).

3.3 Traffic Model

The medical data is collected and summarized as a report by the sensor and transmitted to the smartphone at the end of each time slot. A report contains health information that is required for rapid response [4]. We denote the number of packets in a report N_p. The transmission of reports from sensor to smartphone follows the sum of the Bernoulli process, and the arrival rate for the ith WBAN is given by λ_i. Because condition of a patient changes much slower than the channel variations, we only consider a scenario where $\lambda_i < 1$. The computing capability of sensors is limited; therefore, we assume that the sensor buffer can only store a limited number of packets. With a loss of generality, we consider that a sensor can only cache one report within each time slot. If a new report arrives but the previous report is still cached in the buffer, the previous report is evicted from the buffer and replaced by the new report. The buffer state of the ith WBAN at the beginning of time slot n is given by $q_i(n) \in \{0,1\}$, where $q_i(n) = 0$ means that the sensor buffer of the ith WBAN is empty at the beginning of timeslot n; otherwise, $q_i(n) = 1$.

3.4 Channel-Access Scheme

The goal of MAC is to maximize network throughput without allowing the sensors to consume too much energy. The channel-access scheme is described as follows. At the beginning of each time slot, the central controller sends out a beacon to choose a WBAN for transmission during this time slot. Because only one WBAN is scheduled, interference between WBANs is avoided. Let $s(n)$ denote the index of the WBAN chosen during time slot n, and $s(n) = i$. The smartphone of the ith WBAN sends a beacon to the sensor in the same WBAN. If CM3 is on and there is one report to be transmitted, i.e., $C_i(n) = 1$ and $q_i(n) = 1$, the transmission from the sensor to smartphone will be successful and the sensor's buffer will be emptied. Then the smartphone forwards the report to the central controller; otherwise, only the channel state is reported to the central controller for future scheduling.

There are two pieces of information about network state available to the central controller. One piece of information is the statistics about the random processes of CM3 channels and the medical report event arrival. In practice, the central controller can obtain these statistics by learning over a period of time. Specifically, the smartphone in each WBAN can learn the channel statistics of that WBAN first and forward them to the central controller. Adaptive learning algorithms [24], [3] could be used to increase learning accuracy in real time. We assume that the central controller can obtain accurate information about the body area channels. In the future, we will consider the impact of imperfect and delayed information on MAC design. The other piece of information is the partial real-time information about the network state. As described in the channel-access scheme, the central controller has the information about the WBAN it chooses at the end of each time slot. Thus, real-time information about the network state is partly available.

To make a proper decision at each time slot, the central controller maintains the belief states of the channel and buffer states of all WBANs based on both statistical information and partial real-time information. Let $\Omega(n) \triangleq [\omega_1(n), \cdots, \omega_{N_W}(n)]$ denote the belief states of the channel states of all WBANs at the beginning of time slot n, where $\omega_i(n)$ is the belief state of the channel state of the ith WBAN over time slot n. The belief states evolve as follows. If real-time information of the ith WBAN is available, the central controller updates its belief state of the ith WBAN according to real information; otherwise, the central controller updates this belief according to statistical information. The belief state evolution can be written as

$$\omega_i^c(n+1) = \begin{cases} 1 - b_i, & S(n) = i, C_i(n) = 1; \\ g_i, & S(n) = i, C_i(n) = 0; \\ T_i^c(\gamma), & S(n) \neq i \end{cases} \tag{2}$$

where $T_i^c(\gamma)$ is an evolution operator of the belief channel state of the ith WBAN. For the ON-OFF channel model, the operator is

$$T_i^c(\gamma) = \gamma(1 - b_i) + (1 - \gamma) \tag{3}$$

As described in [16], the above belief state is a sufficient statistic that depicts current channel state given the channel state is a Markov process.

4 Problem Formulation

In this section, we formulate the problem as a partly observable optimization problem. Then, we investigate the value function of the proposed problem for policy design.

4.1 Reward and Objectives

We first design a reward to facilitate the decision-making of the central controller. The reward should favor higher throughput and not favor packet drop. If the ith WBAN is chosen and transmission is successful, $B_i(n)$ units of reward are received by the network. Let $R_i(n)$ denote the reward obtained in time slot n when the ith WBAN is chosen. $R_i(n)$ is given by

$$R_i(n) = C_i(n) q_i(n) B_i(n)\,. \tag{4}$$

If the channel of the chosen WBAN is off or the buffer of the chosen WBAN is empty, the reward is zero; otherwise, $B_i(n)$ amount of reward accumulated. Let $B_i(n)$ equal the probability of one medical report arriving since the last successful transmission. If a WBAN is not given channel access for a long time, the reward that can be received by the WBAN is small because many packets may have been lost.

The issue for the central controller is which WBAN should be given channel access at each time slot. A control policy for this problem is π: $\Omega(n) \rightarrow s(n)$, a function that maps the belief state $\Omega(n)$ to the action $s(n)$. The goal of the central controller is to maximize the average reward of the network over infinite horizon, which is a common measure in communication system [18]. Thus, the control problem can be written as

$$\boldsymbol{P1} \ \max_{\pi} \ E[\lim_{K\to\infty} \sum_{j=1}^{K} R_{\pi(\Omega(j))}(j) | \ \Omega(1)]\,. \tag{5}$$

Let π^* denote the optimal solution to P1, then

$$\pi^* = \arg\ \max E[\lim_{K\to\infty} \sum_{j=1}^{K} R_{\pi(\Omega(j))}(j) | \ \Omega(1)] \tag{6}$$

Because the real-time state of the full network is not observable, P1 is a partly observable optimization problem. If only the channel state is considered, P1 becomes a partly observable Markov decision process problem (POMDP) [16]. A POMDP has larger state space compared with an observable optimization problem and is more difficult to solve. Our problem is more difficult than a POMDP because we consider random traffic arrival. P1 is a dynamic programming problem; thus, we study the value function of P1 for policy design.

4.2 Value Function

Value function analysis involves breaking an optimization problem over multiple periods into sub-problems at different points in time. The value function at time slot n is given by $Vn(\Omega(n))$. This is the maximum expected reward that the network can have at time slot n. We consider a case where the central controller chooses the ith WBAN at the beginning of time slot n and updates the network state information at the end of time slot n. The reward that can be obtained from time slot n comprises the expected immediate reward $E[R_i(n)]$ and the maximum expected reward from timeslot $n+1$, namely V_{n+1} $(\Omega(n+1)|s(n)=i,\ C_i(n))$. Thus, the value function of P1 at time slot n can be written as

$$V_n(\Omega(n)) = \max_{s(n)} \{E[R_i(n)] + \omega_i(n) V_{n+1}(\Omega(n+1)|i,1) + (1 - \omega_i(n)) V_{n+1}(\Omega(n+1)|i,0)\}. \tag{7}$$

For a POMDP, the value function is piecewise linear and convex [17]. However, for a general partly observable problem, the value function may not have this property. Generally, (7) can be solved backwards to obtain the value of $V1(\Omega(1))$ and the optimal policy π^*. However, because computation complexity increases exponentially, the value function and optimal policy cannot be obtained in real time by the central controller.

5 Policy Design

Because obtaining the optimal solution to P1 is difficult, we first investigate the properties of the channel and buffer dynamics and from this investigation we propose a policy.

5.1 System Dynamics

The belief state of the ith WBAN is ω_i at any time slot, and we study what the belief state will be after k consecutive time slots during which the ith WBAN is not chosen by the central controller. Let $(T_i^c(\omega(n)))^k \triangleq \Pr\{C(n+k)=1|\omega(n)\}$ (k= 0 , 1, 2,..., n) denote the belief state evolution for k consecutive unobserved time slots. Then we have [18]

$$(T_i^c(\omega(n)))^k = \frac{g_i}{b_i+g_i} - \frac{(1-b_i-g_i)^k(g_i-(b_i+g_i)\omega(n))}{b_i+g_i}. \quad (8)$$

Fig. 3 shows how $T_i^c(\omega(n))^k$ changes over time given the positive correlation of CM3, i.e., $1-b_i-g_i>0$. The belief state eventually converges to $\Pi_i^c(2)$, which is the stationary probability that the channel is on. This suggests that a policy should work in the following way. If the ith WBAN is chosen and the channel of the ith WBAN is on, the central controller should give preference to this WBAN again in the near future in order to utilize the ON state. In contrast, if the ith WBAN is chosen and the channel of the ith WBAN is off, the central controller should not give preference to this WBAN in near future in order to avoid wasting channel resource.

Second, given the initial state $q_i(n)$ = 0, we study the probability that exactly one report arrives during the duration k consecutive time slots without transmission. This is the reward we set for a successful transmission. Let $N(kT)$ denote the number of arrived reports during a period kT. For a sum of Bernoulli process, the probability of m events arriving during a period kT is $e^{-\lambda kT}(\lambda kT)^m/m!$ Thus, given the initial state $q_i(n)$ = 0, the probability that exactly one report arrives during a duration of k time slots is:

$$\Pr\{N(kT)=1\} = \lambda kTe^{-(-\lambda kT)}. \quad (9)$$

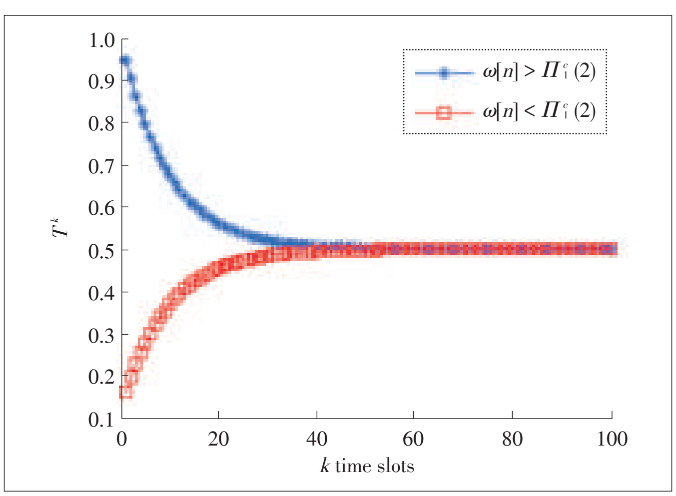
From (9), the probability has a maximum value at 1/λ. Because the system is slotted, the corresponding number of time slot k can be either the minimum integer larger than 1/(λT) or the maximum integer smaller than 1/(λT).

Fig. 4 shows how $\Pr\{N(kT)=1\}$ changes over time. The probability of exactly one report arrival increases to its peak as time goes and then decreases. This result suggests that a control policy should have the following property. If the central controller determines that the buffer of the ith WBAN is empty, it needs to wait for a period of time to revisit the ith WBAN for a new report arrival. However, if the duration is larger than 1/λ, the probability of report loss increases, and this leads to a smaller reward. Thus, the central controller should not wait too long for a revisit.

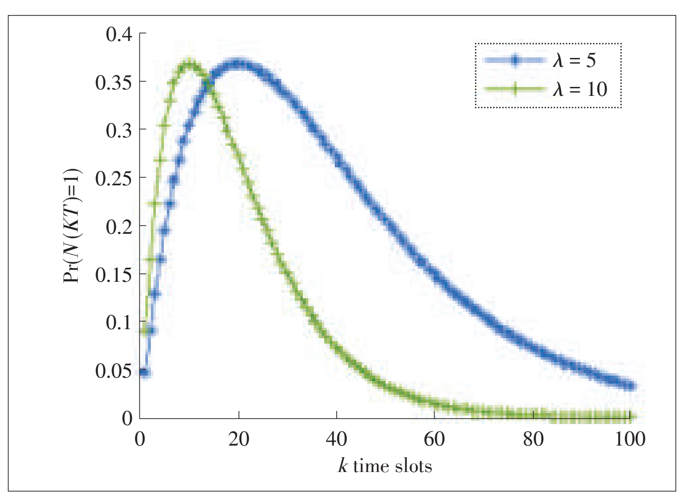
5.2 A Modified Myopic Policy

We construct a myopic policy and, through analysis, point out that this policy results in high report dropping. Drawing on our previous analysis of system dynamics, we propose a modified myopic policy that addresses this issue by approximating the expected future reward.

If only the dynamics of the channel is considered, as in [18], the optimal policy is to stick to the WBAN that is on. Specifi-



▲Figure 3. Evolution of belief state of channel.



▲Figure 4. Evolution of one event arrival probability.

cally, if a WBAN is found to be on, the central controller should keep choosing this WBAN until the channel turns off. With the random report arrivals in WBANs, the above policy is no longer optimal. After a successful transmission, the probability of a report arrival is low (Fig. 4). Thus, even though the channel in the previous slot is on, the central controller does not give preference to this WBAN.

Obtaining an optimal policy is complex; therefore, we develop a myopic policy. The central controller´s objective is simplified to maximize the expected reward for a current time slot based on the belief states and ignores the impact of the current decision on the future reward. Let π^m denote the myopic policy. It can be written as

$$\pi^m = \arg\max E\left[R_{\pi(\Omega(j))}(j)|\Omega(1)\right] \quad (10)$$

The myopic policy π^m has two issues. First, the belief states of the channel converge in a homogeneous network setting (where all WBANs have the same statistics). Therefore, the central controller needs a scheme to choose from multiple WBANs with the same expected rewards. This is not addressed in the myopic policy. Second, with a myopic policy, if a WBAN has not be chosen for more than $1/(\lambda T)$ consecutive time slots, the chance that the central controller will choose this WBAN decreases because the expected reward is smaller. This will cause reports to be dropped in that WBAN. This problem is rooted in the myopic philosophy. An optimal control policy that takes into account future reward does not have such an issue because if a WBAN is not chosen for more than $1/(\lambda T)$ consecutive time slots, the central controller tends to choose this WBAN. Otherwise, the expected future reward is smaller, and this leads to a smaller total reward. In other words, after $1/(\lambda T)$, the myopic policy significantly deviates from the optimal policy.

We propose a modified myopic policy to address the above issues. First, when multiple WBANs have the same maximum expected reward, the central controller chooses a WBAN using a random picker. This random picker makes a choice according to a random number that it generates from a probability density function. In this work, we use a uniform distribution. Second, the impact of future rewards is considered. Because the complexity of obtaining the accurate future reward is high, the future reward is approximated heuristically. Specifically, we increase the expected reward of the current time slot for the WBANs that have been waiting more than $1/(\lambda T)$ time slots:

$$R_i(\tau) = \frac{w\tau\lambda T + 1}{\lambda T} C_i(\tau) q_i(\tau) B_i(\tau) \quad (11)$$

where τ is the number of slot that the ith WBAN has been waiting more than $1/(\lambda T)$, and w is scaling factor. Over time, $(w\tau\lambda T + 1)/\lambda T$ increases. Thus, the WBANs that have been waiting more than $1/(\lambda T)$ time slots have an increased chance of being chosen. This helps solve the second issue introduced by the myopic policy.

6 Simulation Results

In this section, we evaluate the performance of the proposed MAC layer resource allocation through Matlab simulations.

6.1 Simulation Setup

We simulate a scenario similar to that shown in Fig. 1. A central controller is placed at the center of the network, and there are total N_w patients, each with a WBAN for health monitoring. The CM3 channels are simulated using the GE model, and the CM4 channels are simulated as being error-free for packet transmission. The initial channel states of the patients are generated randomly according to the stationary distribution of the channel states. In practice, the arrival rate of vital signs, such as blood pressure and heart rate, is usually less than 100 Kbps whereas the arrival rate of ECG and EMG signals is nearer 1 Mbps [5]. Thus, the network is either congested or uncongested. This leads us to evaluate the effectiveness of the proposed resource allocation policy when the network is congested and uncongested. Given the normalized service capacity, the congested network can be given by

$$\sum_i \lambda_i T > 1 \quad (12)$$

and the uncongested network can be given by

$$\sum_i \lambda_i T < 1 . \quad (13)$$

In the simulation, we vary the report arrival rate to change the network condition. Let λ_c and λ_u denote the report arrival rate for a congested network and uncongested network, respectively. Let N_c^w and N_u^w denote the number of patients in the congested network and uncongested network, respectively.

For simplicity, we consider a homogeneous network where the report arrival rate and the channel statistics of all WBANs are the same. We omit the index of parameters, and the settings are shown in **Table 1**.

In each experiment, we compare our proposal with the round

Table 1. System parameters for simulation

Parameter	Definition	Value
b	Probability that channel turns good	0.15
g	Probability that channel turns bad	0.05
N_w^c	Number of patients in congested network	10
N_w^u	Number of patients in uncongested network	15
λ_u	Arrival rate for uncongested network	5 per second
T	Slot duration	10 ms
λ_c	Arrival rate for congested network	30 per second
w	Scaling factor for modified reward	10

-robin (RR) scheme [25] and myopic (Myo) policy [26]. The RR scheme is chosen because it is simple and starvation-free. In the RR scheme, the central controller assigns channel access opportunity to WBANs in a circular way. The WBAN chosen in time slot n is given by

$$s_{rr}(n) = n \bmod N \tag{14}$$

where mod is the modulo operator. Let $kN \bmod N = N$, for $k \in \{0, 1, 2, \ldots, n\}$ because the network index starts from 1 in this work. The myopic algorithm is chosen because it is optimal when only the channel dynamics are considered [26]. The myopic algorithm is to choose the WBAN with the best belief state of channel.

From here, we call our proposal MyoMo. In each simulation, we report the number of successful transmissions, number of reports dropped, and number of wasted transmission opportunities using our proposal and existing proposals. A transmission opportunity is wasted if a WBAN is chosen but transmission is unsuccessful, either because the channel is OFF or the buffer is empty. The improvements of MyoMo and RR on Myo are expressed as percentage ratio of performance difference to the performance of Myo. Each simulation was 1000 s in duration.

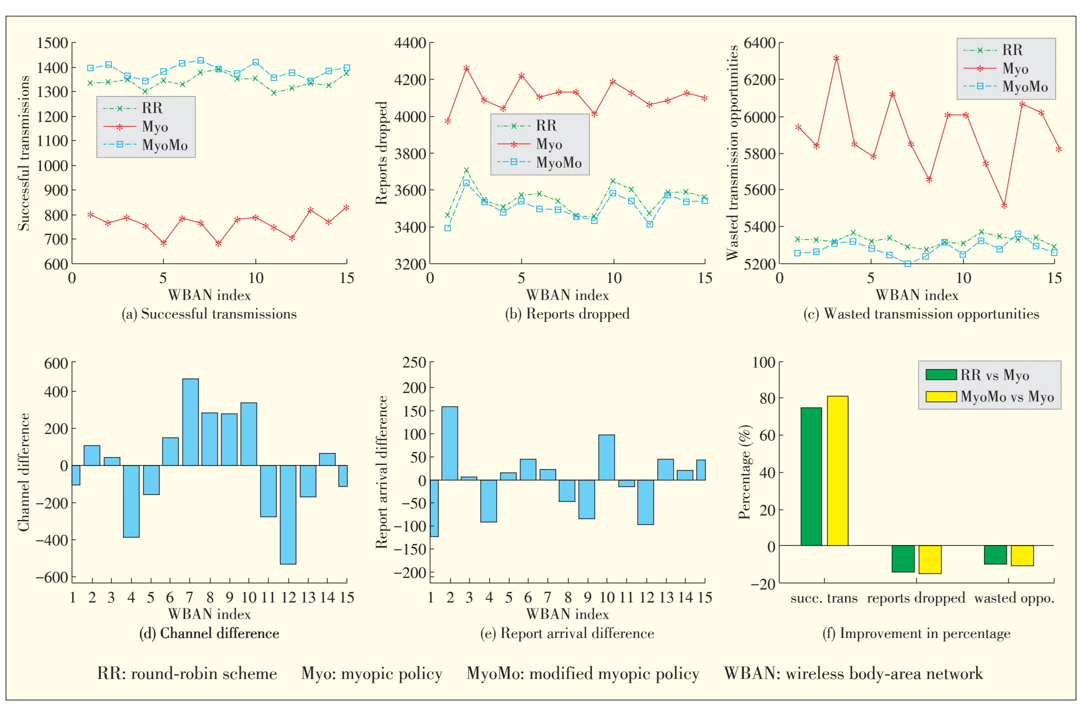
6.2 Performance Evaluation

Here, we report simulation results in terms of 1) network throughput (measured by the number of successful transmissions, 2) the number of reports dropped, and 3) channel utilization (measured by the number of wasted transmission opportunities).

6.2.1 Uncongested Scenario

Fig. 5 shows the performances of RR, Myo and MyoMo algorithms in an uncongested network. Fig. 5a, Myo performs the worst, with about 450 less successful transmissions than RR and MyoMo. MyoMo performs slightly better than RR, with about 50 more successful transmissions on average. Myo causes about 400 more dropped reports than RR and MyoMo, and most WBANs drop fewer reports under MyoMo than under RR (Fig. 5b). The lesser performance of Myo compared to RR and MyoMo can also be seen in Fig. 5c. The superiority of RR and MyoMo over Myo in an uncongested network is shown in Fig. 5f. Compared to Myo, MyoMo and RR have 82% and 77% more successful transmissions, respectively; approximately 15% fewer dropped reports; and approximately 10% fewer wasted transmission opportunities. The reason that Myo performs the worst is that Myo only takes into consideration the channel state and always chooses the WBAN with the best channel. However, the chosen WBAN may have an empty buffer, and this leads to a high number of wasted transmission opportunities. With low channel utilization, the number of reports dropped is high, and the number of successful transmissions is low. The reason that RR performs almost as well as MyoMo is because in an uncongested network, a wasted transmission opportunity is more likely caused by an empty buffer than a channel that is off. The number of WBANs in the network is denoted N_w^u. In the RR algorithm, each WBAN needs to wait for N_w^u time slots for a transmission opportunity. When N_w^u is sufficiently large, the probability of an empty buffer is small. In other words, the RR algorithm aims to avoid an empty buffer. It reduces the number of wasted transmission opportunities and improves the number of successful transmissions. In an uncongested network, the MyoMo algorithm and related reward (11) is dominant because the buffer is taken into account. MyoMo aims to avoid an empty buffer, and this is similar to RR. As a result, the performance of RR is similar to MyoMo when the network is uncongested.

Fig. 5d and Fig. 5e show the difference between WBANs in terms of the number of channels that are on and the number of report arrivals, respectively. In Fig. 5d, the 0 on the y-axis is the average number of channels that are on. The difference in



▲**Figure 5. Performance for unsaturated scenario.**

the number of channels that are on in different WBANs can be more than 1000. The reason for this is twofold: 1) the initial channel states are generated randomly according to the stationary distribution of the channel states, and 2) the channel states evolve according to a probability transition matrix (1). The number of report arrivals is different for different WBANs (Fig. 5e). These differences are the result of the random Poisson number generation method we used in Matlab. The difference between the number of channels that are on and the number of report arrivals causes performance to vary between WBANs using the same algorithm.

6.2.2 Congested Scenario

Fig. 6 shows the performance of the RR, Myo and MyoMo algorithms in a congested network. MyoMo performs the best, with about 300 more successful transmissions than either RR or Myo, which performs the worst. However, the difference in performance between Myo and RR is smaller. These trends can also be seen in report dropping Fig. 6b and channel utilization Fig. 6c. The improvements on Myo brought about by RR and MyoMo are shown in Fig. 6f. Compared to Myo, MyoMo completes 20% more successful transmissions, has 2% fewer report drops, and has 6% fewer wasted transmission opportunities. Compared to Myo, RR completes 8% more successful transmissions, has 1% fewer dropped reports, and has 3% fewer wasted transmission opportunities. MyoMo outperforms RR in congested networks because the number of wasted transmission opportunities caused by empty buffers reduces, and the number of wasted transmission opportunities caused by channels that are off increases. Unlike RR, MyoMo exploits the temporal channel correlation and uses the belief state of the channel to make a decision. Thus, MyoMo is less likely to choose a WBAN with a channel that is off, and this leads to better performance.

From Fig. 6 and Fig. 5, the number of successful transmissions in a congested network is greater than in an uncongested network whereas the number of wasted transmission opportunities in a congested network is smaller than that in an uncongested network. In a congested network, wasted transmission opportunities due to empty buffer are greatly reduced, and this leads to a higher number of successful transmissions.

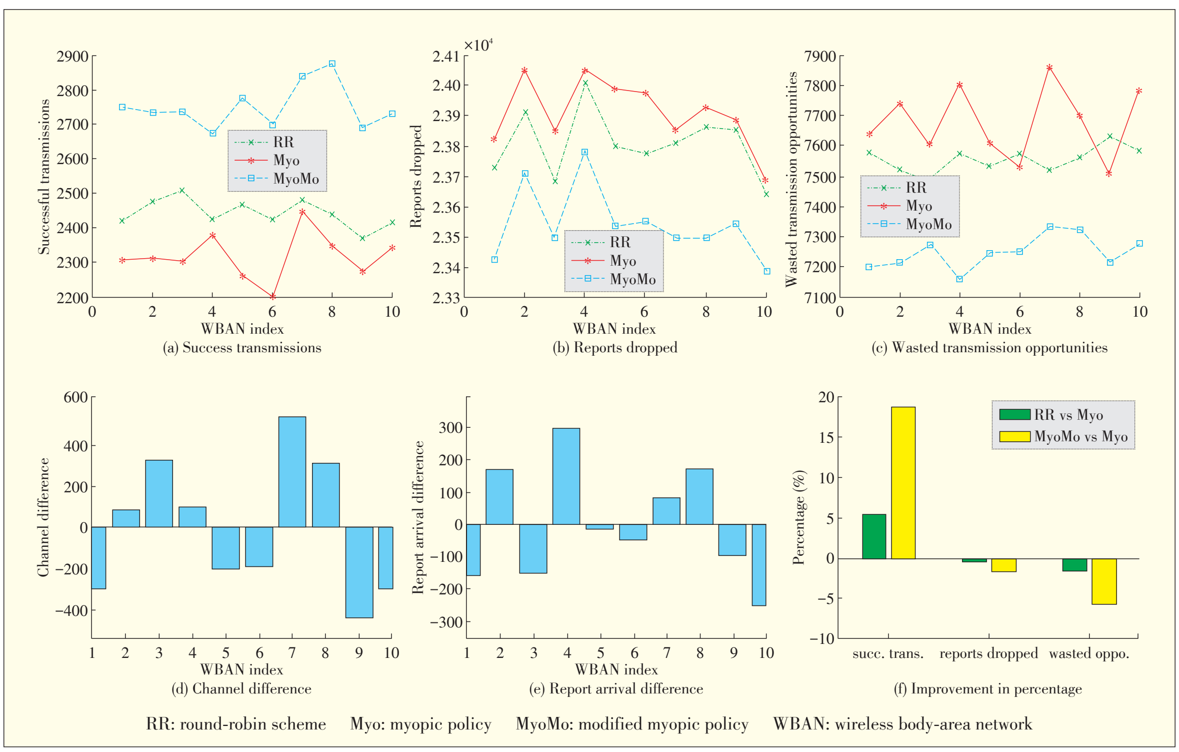


Figure 6. Performance in a congested network.

7 Conclusion and Future Work

We have proposed a MAC layer resource-allocation scheme for a WBAN. Through theory and simulation, we have demonstrated the effectiveness of our proposal in terms of increasing network throughput and channel utilization in both congested and uncongested networks. In future work, we will consider heterogeneous medical data traffic with differential services provisioned using MAC layer resource allocation.

References

[1] Q. Shen, X. Liang, X. Shen, X. Lin, and H. Y. Luo, "Exploiting geo-distributed clouds for a e-health monitoring system with minimum service delay and privacy preservation," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 2, pp. 430-439, 2014. doi: 10.1109/JBHI.2013.2292829.

[2] Q. Shen and W. Zhuang, "Energy efficient scheduling for delay constrained communication in wireless body area networks," in *Proc IEEE GLOBECOM 2012*, Anaheim, USA, pp. 262–267.
[3] X. Liang, X. Li, Q. Shen, R. Lu, X. Lin, X. Shen, and W. Zhuang, "Exploiting prediction to enable secure and reliable routing in wireless body area networks," in *Proc. IEEE INFOCOM 2012*, Orlando, USA, pp. 388–396. doi: 10.1109/INFCOM.2012.6195777.
[4] J. Welch, J. Moon, and S. McCombie, "Early detection of the deteriorating patient: the case for a multi-parameter patient-worn monitor," *Biomedical Instrumentation and Technology*, vol. 46, no. 2, pp. 57–64, 2012.
[5] H. Cao, V. Leung, C. Chow, and H. Chan, "Enabling technologies for wireless body area networks: a survey and outlook," *IEEE Communication Magazine*, vol. 47, no. 12, pp. 84–93, Dec. 2009. doi: 10.1109/MCOM.2009.5350373.
[6] H. Su and X. Zhang, "Battery-dynamics driven TDMA MAC protocols for wireless body-area monitoring networks in healthcare applications," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 4, pp. 424–434, May 2009. doi: 10.1109/JSAC.2009.090507.
[7] B. Otal, L. Alonso, and C. Verikoukis, "Highly reliable energy-saving MAC for wireless body sensor networks in healthcare systems," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 4, pp. 553–565, May 2009. doi: 10.1109/JSAC.2009.090516.
[8] *IEEE Standard for Local and Metropolitan Area Networks Part 15.6: Wireless Body Area Networks*, IEEE Standard 802.15.6, 2012.
[9] K. S. Kwak, S. Ullah, and N. Ullah, "An overview of IEEE 802.15.6 standard," in *Proc. 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies*, Rome, Italy, 2010, pp. 1–6. doi: 10.1109/ISABEL.2010.5702867.
[10] P. Wang and W. Zhuang, "A token-based scheduling scheme for WLANS supporting voice/data traffic and its performance analysis," *IEEE Transactions on Wireless Communications*, vol. 7, no. 5, pp. 1708–1718, May 2008. doi: 10.1109/TWC.2007.060889.
[11] P. Wang, H. Jiang, and W. Zhuang, "A new MAC scheme supporting voice/data traffic in wireless ad hoc networks," *IEEE Transactions on Mobile Computing*, vol. 7, no. 12, pp. 1491–1503, Dec. 2008. doi: 10.1109/TMC.2008.73.
[12] H. Jiang, P. Wang, and W. Zhuang, "A distributed channel access scheme with guaranteed priority and enhanced fairness," *IEEE Transactions on Wireless Communications*, vol. 6, no. 6, pp. 2114–2125, Jun. 2007. doi: 10.1109/TWC.2007.05746.
[13] *Information Technology– Telecommunications and Information Exchange Between Systems Local and Metropolitan Area Networks– Specific Requirements Part 11: Wireless LAN Medium Access Control and Physical Layer Specifications*, IEEE Standard 802.11, 2012.
[14] L. Wang, C. Goursaud, N. Nikaein, L. Cottatellucci, and J. Gorce, "Cooperative scheduling for coexisting body area networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 1, pp. 123–133, Jan. 2013. doi: 10.1109/TWC.2012.120412.112073.
[15] S. Cheng and C. Huang, "Coloring-based inter-WBAN scheduling for mobile wireless body area networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 2, pp. 250–259, Feb. 2013. doi: 10.1109/TPDS.2012.133.
[16] R. D. Smallwood and E. J. Sondik, "The optimal control of partially observable Markov processes over a finite horizon," *INFORMS Operations Research*, vol. 21, no. 5, pp. 1071–1088, 1973.
[17] R. D. Smallwood and E. J. Sondik, "The optimal control of partially observable Markov processes over the infinite horizon: discounted costs," *INFORMS Operations Research*, vol. 26, no. 2, pp. 282–304, 1978.
[18] K. Liu and Q. Zhao, "Indexability of restless bandit problems and optimality of whittle index for dynamic multichannel access," *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5547–5567, Nov. 2010. doi: 10.1109/TIT.2010.2068950.
[19] H. Luo, "Wearable mini-size intelligent healthcare system," US Patent App. 11/675,684, 2007.
[20] K. Y. Yazdandoost and K. Sayrafian-Pour, "Channel model for body area network (ban)," IEEE P802.15-08-0780-09-0006, Apr. 2009.
[21] E. Reusens, W. Joseph, B. Latré, B. Braem, G. Vermeeren, E. Tanghe, L. Martens, I. Moerman, and C. Blondia, "Characterization of on-body communication channel and energy efficient topology design for wireless body area networks," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 6, pp. 933–945, Nov. 2009. doi: 10.1109/TITB.2009.2033054.
[22] S. Rashwand and J. Mišić, "Channel and error modeling for wireless body area networks," *Springer Mobile Networks and Applications*, vol. 19, no. 3, pp. 276–286, Jun. 2013. doi: 10.1007/s11036-013-0449-8.
[23] P. Sadeghi, R. Kennedy, P. Rapajic, and R. Shams, "Finite-state Markov modeling of fading channels — a survey of principles and applications," *IEEE Signal Processing Magazine*, vol. 25, no. 5, pp. 57–80, Sept. 2008. doi: 10.1109/MSP.2008.926683.
[24] J. Gauvain and C. Lee, "Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains," *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 2, pp. 291–298, Apr. 1994. doi: 10.1109/89.279278.
[25] X. Shen, H. Yu, J. Buford, and M. Akon, *Handbook of Peer-to-Peer Networking*. New York: Springer, 2010.
[26] Q. Zhao, B. Krishnamachari, and K. Liu, "On myopic sensing for multi-channel opportunistic access: structure, optimality, and performance," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 5431–5440, Dec. 2008. doi: 10.1109/T-WC.2008.071349.

Manuscript received: 2014-03-31

Biographies

Qinghua Shen (q2shen@uwaterloo.ca) received his BS degree and MS degree in electrical engineering from Harbin Institute of Technology, China, in 2008 and 2010. He is currently working toward his PhD degree in the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include resource allocation for e-healthcare system and cloud computing.

Xuemin (Sherman) Shen (sshen@uwaterloo.ca) received his BSc (1982) degree from Dalian Maritime University, China, in 1982. He received his MSc degree (1987) and PhD degree (1990) in electrical engineering from Rutgers University, USA, in 1987 and 1990. He is professor and university research chair in the Department of Electrical and Computer Engineering, University of Waterloo, Canada. Dr. Shen's research focuses on resource management in interconnected wireless/wired networks, wireless network security, social networks, smart grid, and vehicular ad hoc and sensor networks. He has chaired or co-chaired many committees of international IEEE conferences and is the editor-in-chief of *IEEE Network, Peer-to-Peer Networking and Applications*, and *IET Communications*. He has also been the founding editor, associate editor, or guest editor for many other peer-reviewed publications. Dr. Shen is a registered professional engineer in Ontario, Canada; fellow of the IEEE; fellow of the Engineering Institute of Canada; fellow of the Canadian Academy of Engineering; and distinguished lecturer at the IEEE Vehicular Technology Society and IEEE Communications Society.

Tom H. Luan (tom.luan@deakin.edu.au) received his BE degree from Xi'an Jiaotong University, China, in 2004. He received his MPhil degree from Hong Kong University of Science and Technology, in 2007. He received his PhD degree from the University of Waterloo, Canada, in 2012. He currently lectures on mobile and applications in the School of Information Technology, Deakin University, Australia. His research interests include mobile cloud computing, mobile application and service development, vehicular networking, and wireless content distribution.

Jing Liu (Jingliu_lj@stju.edu.cn) received her BS, MS and PhD degrees from Xidian University, China, in 1998, 2001 and 2005. Since July 2005, she has worked in the Department of Electronic Engineering, Shanghai Jiao Tong University, China. She is currently an associate professor at Shaighai Jiao Tong University. From 2012 to 2013, she was a visiting professor in the Department of Electrical and Computer Engineering, University of Waterloo, Canada. Her current research interests include wireless body area networking, wireless sensor networking, cooperative communication, cognitive communication, and LTE networking.

DOI: 10.3939/j.issn.1673-5188.2014.03.003
http://www.cnki.net/kcms/detail/34.1294.TN.20140829.1544.003.html, published online 29 August, 2014

Selective Cluster-Based Temperature Monitoring System for Homogeneous Wireless Sensor Networks

Sudhanshu Tyagi[1], Sudeep Tanwar[2], Sumit Kumar Gupta[3], Neeraj Kumar[4], and Joel J. P. C. Rodrigues[5, 6]
(1. Department of Electronics and Communication Engineering, J. P. Institute of Enginerring and Technology (JPIET), Meerut 250002, India;
2. Department of Information Technology, Bharat Institute of Technology, Meerut 250002, India;
3. Department of Electronics and Communication Engineering, Institute of Information & Management Technology (IIMT), Meerut 250002, India;
4. Department of Computer Science and Engineering, Thapar University, Patiala 147004, India;
5. Instituto de Telecomunicações, University of Beira Interior, Covilhã 6200-034, Portugal;
6. University ITMO, Saint-Petersburg 197101, Russia)

Abstract

Over the past few decades, there has been a revolution in ICT, and this has led to the evolution of wireless sensor networks (WSN), in particular, wireless body area networks. Such networks comprise a specialized collection of sensor nodes (SNs) that may be deployed randomly in a body area network to collect data from the human body. In a health monitoring system, it may be essential to maintain constant environmental conditions within a specific area in the hospital. In this paper, we propose a temperature-monitoring system and describe a case study of a health-monitoring system for patents critically ill with the same disease and in the same environment. We propose Enhanced LEACH Selective Cluster (E-LEACH-SC) routing protocol for monitoring the temperature of an area in a hospital. We modified existing Selective Cluster LEACH protocol by using a fixed-distance-based threshold to divide the coverage region in two subregions. Direct data transmission and selective cluster-based data transmission approaches were used to provide short-range and long-distance coverage for the collection of data from the body of ill patients. Extensive simulations were run by varying the ratio of node densities of the two subregions in the health-monitoring system. Last Node Alive (LNA), which is a measure of network lifespan, was the parameter for evaluating the performance of the proposed scheme. The simulation results show that the proposed scheme significantly increases network lifespan compared with traditional LEACH and LEACH-SC protocols, which by themselves improve the overall performance of the health-monitoring system.

Keywords

network lifespan; sensor node; LEACH; LNA

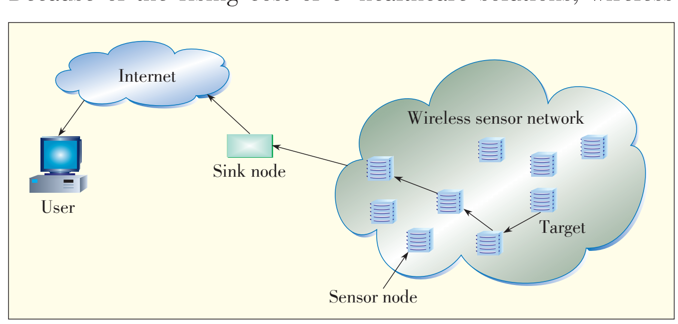
1 Introduction

A wireless sensor network (WSN) comprises a large number of sensor nodes (SNs) that communicate wirelessly. In a WSN, a sink node collects data from all sensors for further processing. A WSN is considered a multihop network for cluster communication. **Fig. 1** shows the schematic of various operations in a WSN.

Each SN has a radio transceiver with internal antenna, a microcontroller, a battery, and an electronic circuit for interfacing with sensors and energy source. Size and cost constrains SNs in terms of their energy consumption (a key issue), memory, computational speed, and communication bandwidth.

The technique used to transmit data from source to sink node is based on an appropriate routing decision, but it is difficult to select a routing scheme is energy efficient.

There are a number of protocols for routing in a WSN. These protocols vary from application to application and may be data-centric, hierarchical, or based on cluster, location, mobility, QoS, network flow, multiple paths, heterogeneity, or homogeneity [1], [2]. Energy efficiency is the key parameter for WSN routing protocols because the battery power in SNs is limited. Because of the rising cost of e-healthcare solutions, wireless



▲Figure 1. Wireless sensor network.

body area networking (WBAN) has received more attention from researchers and application developers [3]. A WBAN is a network of small, intelligent sensors that are attached to clothing, stuck directly onto the body, or implanted inside the body for continuous monitoring. These devices can sense, sample, and send the captured data to an associated device, e.g., a smartphone. Then, this data can be sent via wireless link to a medical team for real-time diagnosis. In a body sensor network, sensors and actuators are used. A sensor detects certain vital signs from the human body, such as blood pressure, heartbeat, and body temperature. An actuator performs specific tasks according to data received from sensors or through interaction with a human. The tiny sensors used in a WBAN are energy-constrained, so energy efficiency is an important research area. We consider a case where the temperature within a hospital is monitored and controlled for patients with similar diseases. The literature shows that clustering is a better mechanism for measurement of unique data. The cluster-based routing protocol is one of the most interesting research areas in WSN. Low-Energy Adaptive Clustering Hierarchy (LEACH) [4] protocol is energy efficient and aggregates data, so it is widely used in WSNs. In this paper, we refer to existing work in [4] and [5] and propose Enhanced Low-Energy Adaptive-Cluster-Based Hierarchical Selective Clustering (E-LEACHSC) protocol. This protocol leverages direct communication and cluster-based communication, both of which occur over the selective cluster. The authors have used fixed distance between BS and CH for communications. From the literature survey in [2], a clustering approach involves a minimum of two hops in the transmission of data from source to base station (BS), even though the distance between the source and BS may be very small. Hence, we focus on the energy constraints on an individual SN. The key parameter use to evaluate our proposed protocol is last node alive (LNA).

The remainder of this paper is organized as follows. In section 2, related work and contributions are discussed. In section 3, E-LEACH-SC routing protocol in described. In section 4, our protocol is simulated in MATLAB, and the performance of our protocol is compared with that of LEACH and LEACH-SC, other well-known routing protocols. Section 5 concludes the paper.

2. Related Work and Contributions

2.1 Related Work

WSN has already been used in a wide variety of applications, so it is important to establish specialized methodoligies for increasing network lifespan and improving routing. A number of proposals have been made to address these issues.

2.1.1 Energy-Harvesting Approach

To improve the lifespan of a WSN, extra energy is required. There are different mechanisms by which the overall energy of the network field can be increased. In this section, we consider the concept of harvesting extra energy from the surrounding environment. Pei et al. [6] suggest a WSN routing protocol based on local adaptive sampling. The authors harvest energy to improve the overall network performance. Adaptive sampling is used to adapt the sample mode to current conditions and to monitor energy in real time. Zhang et al. [7] proposed an optimization algorithm for single clusters and multiple clusters. This algorithm considers the energy-harvesting (EH) node as the relay node. The authors determined the weight of the EH node in order to estimate an optimal CH position. Designing an EH node is not easy, and several things need to be presumed. In [6], the type of EH node used in the proposed work is not discussed.

2.1.2 Application-Oriented Approach

The rising cost of healthcare has shifted the focus towards wireless body area networking (WBAN), which is an emerging area of networking. We have extended the concept of healthcare monitoring and WBAN to provide e-healthcare solutions at cheap rates to end users without compromising quality. J. M. L. P. Caldeira et al. [8] propose using biosensors to monitor temperature within the body. A biosensor is an analytical device used for detection in different applications, and it can vary in sensitivity. The authors continuously measured the intra-vaginal temperature of hospitalized women and found that the temperature varied for each woman. The study had implications for preventing preterm labor, detecting pregnancy contractions, monitoring ovulation period, determining the effectiveness of gynecological treatments, and discovering new contraception methods. Ousmane Diallo et al. [9] propose a framework that combines WBAN with cloud computing and statistical modeling in order to maximize energy efficiency and securely store data received by the sink from body sensors. The authors assume a fixed WBAN topology where the patient in a hospital bed or at home and not in a condition to move. The data of every patient is stored in a server at the hospital so that any queries about the patient can be answered. If the patient's data cannot be found or is insufficient, the query is forwarded to the WBAN so that the required data can be obtained. This scheme enables the retrieval of data that is suitable for real-time medical diagnosis and energy-efficient processing. The authors of [9] do not take into consideration patient mobility. O. R. E. Pereira et al. [10] propose a BSN mobile solution for biofeedback monitoring on Symbian, Windows Mobile, Android, and iPhone over SHIMMER platforms. The authors found that the life of a critically ill person can be improved using a mobile healthcare system as opposed to a traditional healthcare system. In a mobile healthcare system, data can be collected, visualize, and monitored immediately by a team of doctors, and instant action can be taken. Energy efficiency is significant issue in mobile healthcare system and needs to be carefully considered to improve the overall quality of the system.

2.1.3 Direct Communication

Intanagonwiwat et al. [11] proposed directed diffusion (DD) as a way of constructing the route between SNs and BS. There are four stages in DD, the most important being direct data communication. In DD, SNs that are far from the base station die quickly because of the large transmission distance. This means that some areas have no coverage. However, for small-scale networks, DD is a good option.

2.1.4 Cluster-Based Homogeneous Network

Heinzelman et al. [4] proposed a low-energy adaptive clustering hierarchy (LEACH) protocol. This protocol is a self-organizing, adaptive clustering protocol that provides each SN with equal energy. The LEACH protocol is executed in rounds, each of which has two phases: setup and steady state. The steady-state phase is extensive compared to the setup phase. In LEACH protocol, the SNs organize themselves into local clusters, and one node, called the cluster head (CH), is the responsible node. The rest of the nodes, called cluster members (CMs) act as ordinary nodes for the respective cluster. To prolong the life of the network, LEACH protocol randomly rotates the high-energy CH and performs local data fusion to transmit the data from the CHs to the BS. If the BS is far from the network, the energy of the CHs is affected because only CHs communicate directly with the BS. This unique feature of LEACH protocol saves overall network energy because CMs do not directly communicate with the BS.

For a cluster formation, each SN decides whether or not to become a CH for the current round. This decision is based on the percentage of CHs for the network and the number of times the SN has already been a CH. In this process, every SN has to select a random number between 0 and 1. If the number is less than a threshold, the SN becomes a CH for the current round. The threshold is given by [4]:

$$T(n)=\begin{cases}\dfrac{p}{1-p(r \bmod \frac{1}{p})}, n\in G\\ 0,\ \text{otherwise}\end{cases} \tag{1}$$

where p is the desired percentage of CHs, r is the current round, and G is the set of SNs that have not been CHs in the last $1/p$ rounds. During round 0, every SN has a probability p of becoming a CH. The SNs that are CHs in round 0 cannot be CHs for the next $1/p$ rounds. This is necessary to increase the probability of the remaining SNs becoming CHs because now there are fewer SNs that are eligible to become CHs. Using this threshold, each SN will be a CH at some point within $1/p$ rounds. The main drawback of LEACH protocol is that the energy cost is directly proportional to communication distance, i.e., the energy cost increases as the communication distance increases. Therefore, selecting the CMs on the basis of received signal strength increases the communication cost of nodes, which leads to a decrease in energy efficiency across the network.

Heinzelman et al. [12] proposed a centralized, low-energy adaptive clustering hierarchy (LEACH-C) protocol. The main difference between LEACH and LEACH-C is that LEACH does not take into account the remaining energy of SNs whereas LEACH-C does. In LEACH-C protocol, energy-efficient clustering and routing are combined for application-specific data aggregation. This increases network lifespan and decreases data access latency. LEACH-C includes a new distributed cluster formation technique that enables self-organization for a large numbers of SNs.

In CH election, each node elects itself as the CH at beginning of time t with probability p. If the expected number of CHs for this round is assumed to be such that N is the total number of nodes in the network, then $\sum_{i=1}^{N} p_i \times 1 = n$.

The authors of [12] give a mathematical formula for estimating the probability that a node becomes a CH after finite number of iterations:

$$p_i(t)=\begin{cases}\dfrac{n}{N-n\times\left(k \bmod \frac{n}{N}\right)}, f(t)=1\\ \end{cases} \tag{2}$$

where $f(t)$ is the parameter used for the estimation. If the node has been a CH in the most recent rounds then $f(t)$ = 1; otherwise, $f(t)$ = 0. After n/N rounds, each node is expected to become a CH. From (2), the SNs that have not become CHs in recent rounds and have more residual energy than other SNs will become CHs in subsequent rounds. When all the SNs have equal initial energy, the node having minimum residual energy is elected as a CH in the subsequent phases. The remaining execution is same as that of LEACH [4].

Handy et al. [13] proposed a low-energy adaptive clustering hierarchy with deterministic cluster head selection (LEACH-DCHS) protocol. This protocol is an extended version of LEACH protocol. In LEACH-DCHS, the CH is selected by a deterministic component rather than stochastic component. The operation of LEACH-DCHS is similar to that of LEACH, where execution is divided into CH election and cluster formation. However, the threshold for CH selection is modified and is given by

$$T(n)=\begin{cases}\dfrac{p}{1-p\left(r \bmod \frac{1}{p}\right)}\left[\dfrac{E_{n-\text{current}}}{E_{n-\max}}\right], n\in G\\ 0,\ \text{otherwise}\end{cases} \tag{3}$$

This threshold is further modified because it is a low-value threshold, and it is possible that, after some rounds, the network may jam. This modified threshold is given by

$$T(n)=\begin{cases}\dfrac{p}{1-p\left(r \bmod \dfrac{1}{p}\right)}\left[\dfrac{E_{n-\text{current}}}{E_{n-\max}}+\left(r,\text{div}\dfrac{1}{p}\right)\left(1-\dfrac{E_{n-\text{current}}}{E_{n-\max}}\right)\right], n\in G\\ 0, \text{ otherwise}\end{cases} \quad (4)$$

The advantage of LEACH-DCHS is that it improves network lifespan. However, a major drawback of the three protocols discussed here is that the respective CMs are selected on the basis of the strength of the signal received from the CHs.

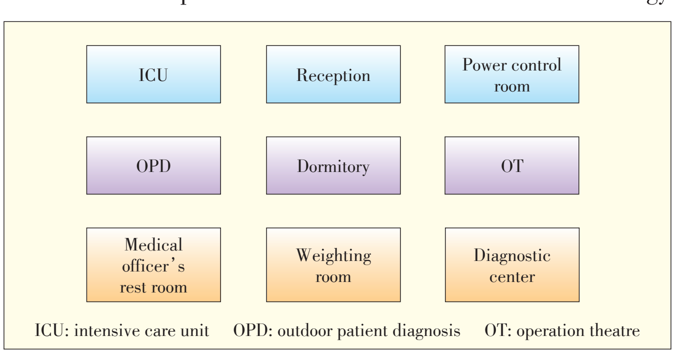
To address the issues of LEACH (and its related protocols) and to prolong network lifespan, Jun et al. [7] propose a low-energy adaptive-clustering hierarchical selective cluster (LEACH-SC) protocol. In LEACH-SC, the process of joining the nodes with the cluster is changed, i.e., the CH closest to the node itself and the sink node is selected. This increases the lifespan of the network. To minimize the amount of energy used by the network and prolong the network lifespan, the authors of [6] use minimum transmission energy, given by

$$Min\left(E_{TX}(k,d)\right)=E_{elec}(k)+E_{TX-amp}(k,d) \quad (5)$$

where k is the number of bits forwarded over distance d, and E_{TX} is the transmitter circuitry dissipation per bit. The distance d is vital to the overall energy cost of the network. **Fig. 2** shows the generalized architecture used in the proposed scheme.

2.2 Contribution

In [14]–[17], LEACH and its related protocols were very good cluster-based protocols; however, when the BS is located at one corner of the network, all the nodes have better load distribution through random selection of CH. SNs opposite the BS die quickly compared to SNs near the BS. We propose Enhanced Low-Energy Adaptive Selective Cluster Based Hierarchical (E-LEACH-SC) protocol. This protocol has the advantage of both direct communication and cluster-based communication. A fixed-distance threshold splits the coverage region into two parts. Results show that E-LEACH-SC considerably increases the lifespan of the network. The network and energy models used for E-LEACH-SC are same as those in [3] and [6]; the only difference is that the BS in our model is located in a 100 × 100 m network.



▲Figure 2. Hospital layout.

3 Proposed Approach

In a small network, the DD approach is always better; however, in a medium to large network field, a cluster-based approach is better. E-LEACH-SC protocol leverages both approaches. E-LEACH-SC used a fixed-distance-based threshold that divides the network field into two segments. The BS is assumed to be located at one of the corners of network and is fixed. Therefore, the direct-communication approach is used for the region near the BS, and the selective cluster-based approach is used for the region further away from the BS (**Fig. 3**).

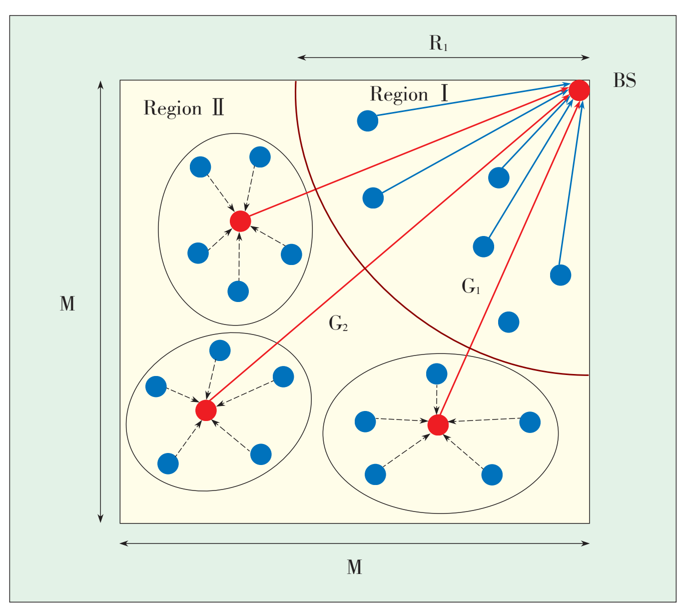
An important issue is how to calculate the fixed-distance-based threshold. We select the appropriate ratio of node densities of the two regions. In LEACH [4] and LEACH-SC [7] protocols, it is assumed that almost no energy is consumed in the formation of cluster. However, in practice, this is not possible. We combine the DD [11] and LEACH-SC [7] approaches, together and using the fixed-distance-based threshold, we save the energy of the region over which clustering is not being used. As a consequence, the overall lifespan of the network is increased.

Theorem 1: The energy of an SN can be used efficiently if transmission distance is kept to a minimum.

Proof 1: For proof of Theorem1, we prove the following lemmas.

Lemma 1: SN is part of a cluster if CH is near the midpoint between an SN and BS.

Proof 1: LEACH is the fundamental cluster-based hierarchical routing protocol. CMs are selected according to the



▲Figure 3. Network model.

strength of the signal received from different CHs. Jun et al. [7] use a different mechanism for selecting CMs [7]. From the analysis in [7], lemma 1 can be proved.

Lemma 2: The cost of communication is lower for a selective cluster.

Proof 2: With reference to lemma 1, the distance between the CH and SN is shorter for a LEACH-SC than LEACH. The distance between CH and the ith non-CH node in a selective cluster is denoted d_{i-sc}, and the distance between CH and the ith non-CH node in an ordinary cluster is denoted d_{i-oc}. If d_{CH-BS} is the distance between CH and BS and there are m CMs in a particular region, then average distance for a selective cluster will be

$$D_{AVG-SC} = \frac{\sum_{i=1}^{m} d_{i-SC}}{m} \tag{6}$$

and the average distance for an ordinary cluster will be

$$D_{AVG-OC} = \frac{\sum_{i=1}^{m} d_{i-OC}}{m}. \tag{7}$$

Therefore, per-round energy consumption of a CH in a selective cluster is

$$E_{r-SC} = (m+1)\times k\times E_{elec} + m\times k\times \varepsilon_{amp}\times D^2_{Avg-SC} + k\times E_{DA} + k\times \varepsilon_{amp}\times D^2_{SC-BS} \tag{8}$$

and per-round energy consumption of a CH in an ordinary cluster is

$$E_{r-OC} = (m+1)\times k\times E_{elec} + m\times k\times \varepsilon_{amp}\times D^2_{Avg-OC} + k\times E_{DA} + k\times \varepsilon_{amp}\times D^2_{SC-BS.} \tag{9}$$

We observe that $D_{Avg-OC} > D_{Avg-SC}$; therefore, lemma 2 is proved, and $E_{r-SC} < E_{r-OC}$.

Lemma 3: A combinational approach is more energy efficient than an individual approach.

Proof 3: For the proof of lemma 3, refer to the theorem in [18]. By considering all lemmas simultaneously, Theorem 1 can be proved.

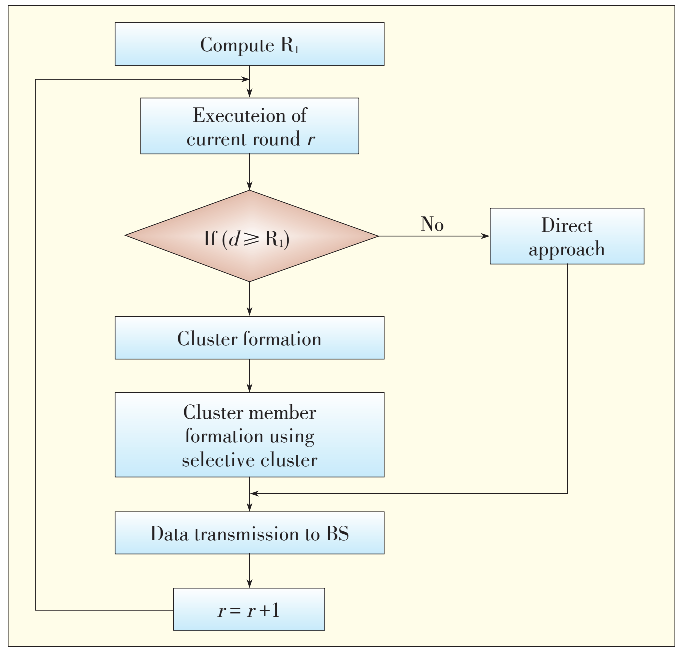
3.1 Execution of E-LEACH-SC

The energy needed in cluster-based communication is higher than that needed for direct communication if the distance between the SN and BS is sufficiently smaller than the distance between the SN and CH and the distance between the CH and BS. **Fig. 4** shows the round-wise execution of our scheme.

3.2 Pseudo Code of E-LEACH-SC

Algorithm 1 shows the pseudo code for our proposed scheme.

All input parameters are initialized in the very first step. On selected node densities, partitions can be formed using an estimated fixed-distance-based threshold. We have already assumed that the network model is homogeneous; hence, all SNs are distributed uniformly over the area of investigation. The SNs of region one communicates directly with the BS. In region



▲**Figure 4. Energy mode flow of E-LEACH-SCl.**

Algorithm 1. E-LEACH-SC: Enhanced LEACH Selective Cluster

```
1: Initialize the parameters as N = TotalSNs,
   N^M = CM, N^B = BS
   d1 = distance from midpoint of SN-BS to CH
   d2 = distance from SN to CH
2: for i = 1; I ≤ N; i++ do
3:   S(i). E = E0
4:   if (d ≤ R1) then
5:     Use DD [7] technique
6:   else
7:     Go to line 8
8:     Elect the CHs based upon T(n)s
9:     Elected CH broadcast the message to all the SNs]
10:  end if
11: end for
12: for ( j = 1; i ≤ N^H ; j ++) do
13:   for ( j = 1; i ≤ N ; j ++) do
14:   if d^2_1(i) < d^2_2(i) then
15:       S(i) = N^M
16:     end if
17:   end for
18:   for (k = 1; k ≤ N^M; k ++) do
19:       CH allocate TDMA schedule to CMs
20:       Data transmitted from CMs to CH as per TDMA schedule
21:       Data aggregated by CH
22:   end for
23:   Data transmitted from CHs to BS
24: end for
```

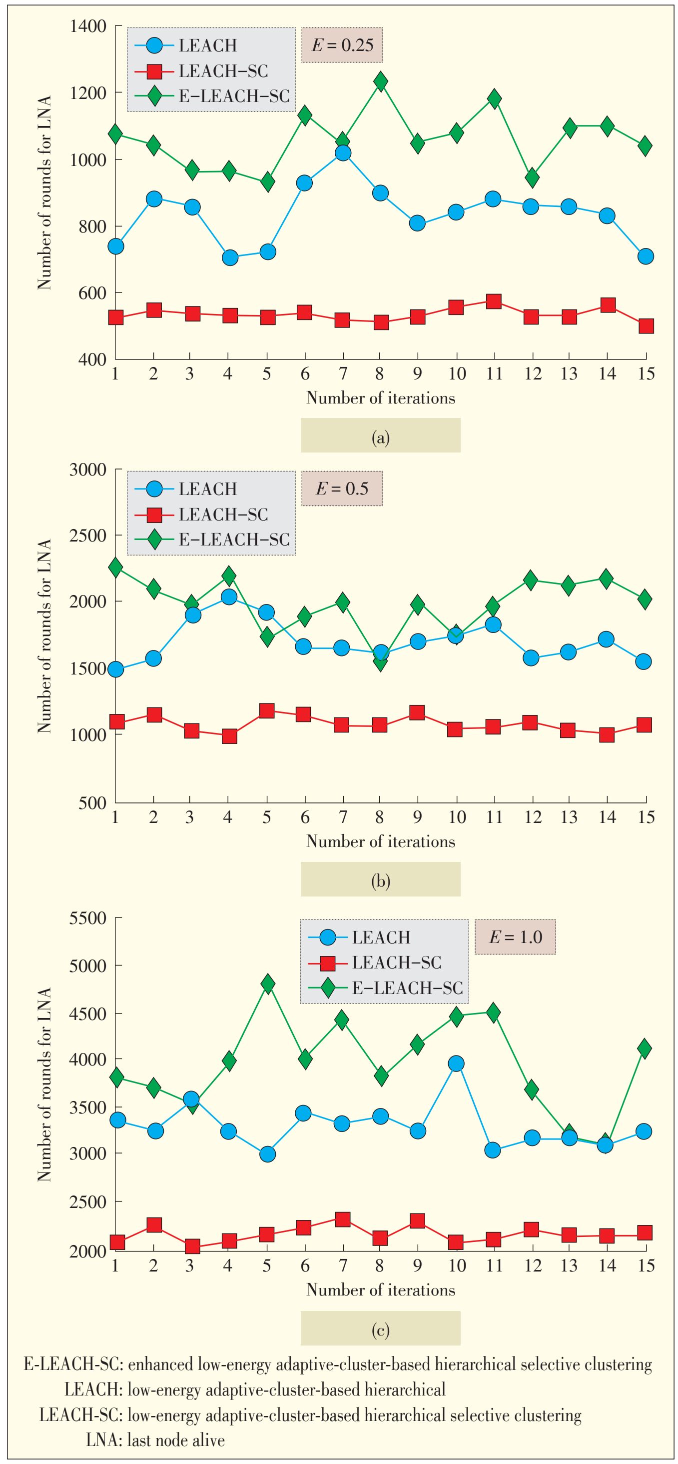
two, selective-cluster-based communication is used (Algorithm 1, steps 11 to 23).

4 Performance Evaluation

We use a first-order radio model to simulate LEACH [4], LEACH-SC [7] and E-LEACH-SC. The parameters used for this simulation are the same as those in [4], [7]. To determine the performance of the E-LEACH-SC protocol, we had simulated a homogeneous clustered WSN in a 100 × 100 m^2 field. The total number of SNs is 100, and these SNs are uniformly distributed over the network. All the SNs are stationary. The message for data packet transmission is 4000 bits. The key parameters for determining the performance of the E-LEACH-SC protocol are network lifespan increase and system stability. LNA is used to observe the abovementioned issues for different initial energy values. **Fig. 5** shows that the E-LEACH-SC protocol significantly improves system lifespan compared to the LEACH and LEACH-SC protocols.

Fig. 6 shows the node densities of two regions with different initial energy levels. The results validate the selection of a fixed-distance-based threshold for splitting the network into two regions. In this case, we also use LNA as the measuring parameter.

Fig. 7 shows network stability for LEACH, LEACH-SC and E-LEACH-SC protocols for different initial energy levels. The average number of rounds for fifteen iterations in LEACH and LEACH-SC is less than that for E-LEACH-SC. Looking at LNA performance, we observe that our proposed protocol is more stable than the LEACH and LEACH-SC protocols.



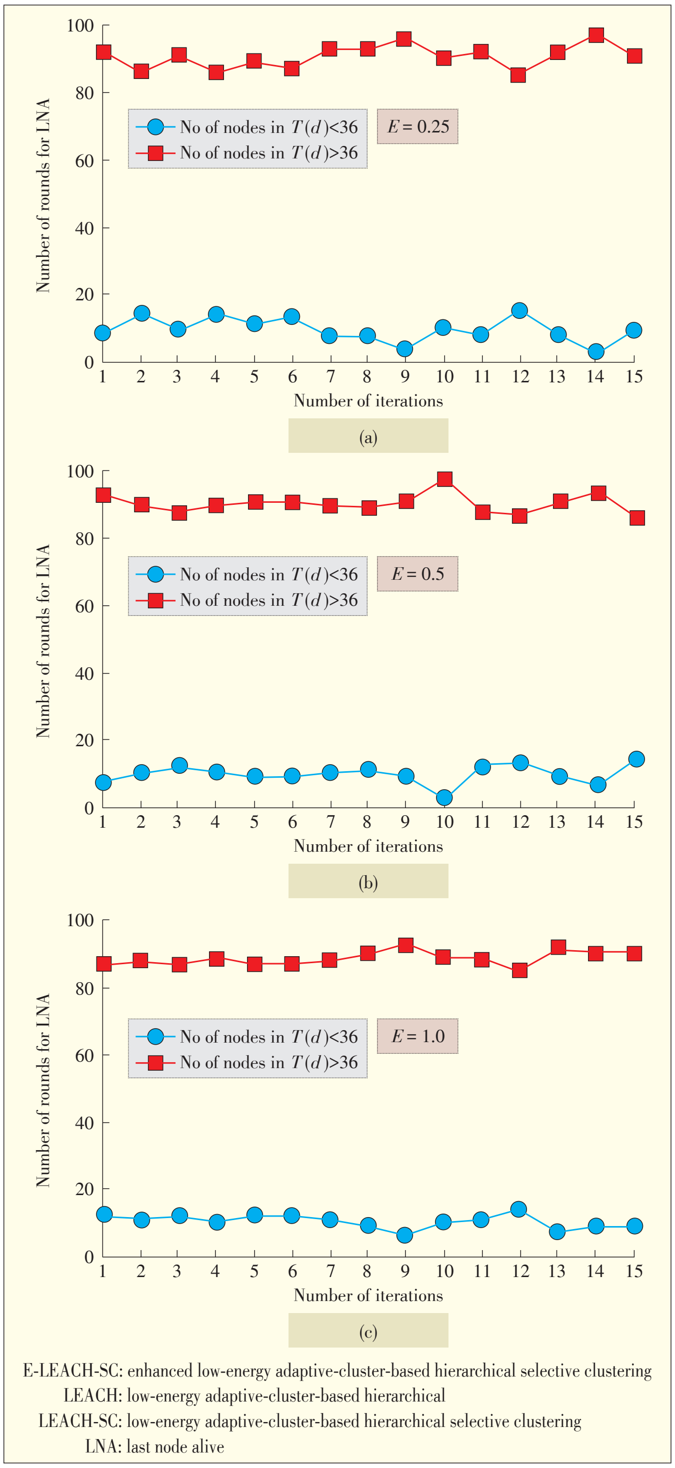
▲**Figure 5. Network lifespan as LNA for LEACH, LEACH-SC and E-LEACH-SC at initial energy levels a) E_0 = 0.25J , b) E_0 =0.5J , and c)E_0 =1.0J.**

5 Conclusion

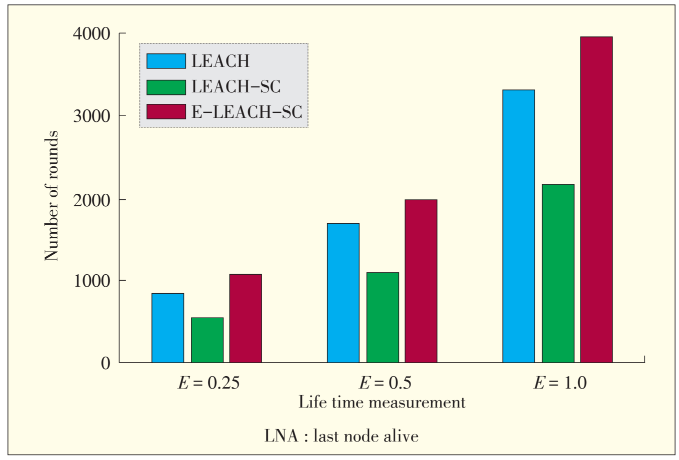
WSNs have been used in a large number of medical and healthcare applications. Increasing the lifespan of SNs increases the lifespan and stability of the entire WSN. In this paper, we have proposed Enhanced Low-Energy Adaptive Selective-Cluster-based Hierarchical (E-LEACH-SC) protocol for WSN. The E-LEACH-SC protocol increases system stability, is easy to implement, and extends the lifespan of the network. Our protocol uses selective-cluster-based communication as well as direct communication between nodes and BS on the basis of a fixed-distance-based threshold. Simulations show that E-LEACH-SC increase the lifespan of the network and provides good stability compared with LEACH and LEACH-SC for the same energy level. In the future, we intend to compare E-LEACH-SC with other protocols and using additional parameters. Security aspects of the current scheme will also be explored.

Acknowledgements

This work has been partially supported by Instituto de Telecomunicações, Next Generation Networks and Applications Group (NetGNA), Covilhã Delegation, by Government of Russian Federation, Grant 074-U01, and by National Funding from the FCT - Fundação para a Ciência e Tecnologia through the Pest-OE/EEI/LA0008/2013 Project.



▲Figure 6. Node densities of two regions for ELEACH-SC at initial energy levels a) E_0 = 0.25 J, b) E_0 = 0.5 J, and c) E_0 = 1.0 J.



▲Figure 7. Stability of LEACH, LEACH-SC and E-LEACH-SC protocols for different initial energy levels.

References

[1] W. L. Su, S. Y. Sankara, and E. Cayirci. "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102-114, 2002. doi: 10.1109/MCOM.2002.1024422.

[2] D. Cook and D. Sajal, *Smart Environments: Technologies, Protocols, and Applications*. England: John Wiley&Sons, pp. 56-63, 2004.

[3] D. B. He, N. Kumar, J. H. Chen, C. C. Lee, N. Chilamkurti, and S. S. Yeo, "Robust anonymous authentication protocol for health-care applications using wireless medical sensor networks," *Multimedia Systems*, In Press, 2014. doi: 10.1007/s00530-013-0346-9.

[4] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan. "Energy-efficient communication protocol for wireless microsensor networks," in *Proc. IEEE Hawaii Inter. Conf. on System Sciences*, Maui, HI, USA, 2000, pp. 110. doi: 10.1109/HICSS.2000.926982.

[5] W. Jun, Z. Xin, J. Y. Xie, and Z. K. Mi, "A Distance-based Clustering Routing Protocol in Wireless Sensor Networks," in *Proc. 2010 12th IEEE Int. Conf. on Commun. Technology*, Nanjing, China, Nov. 2010. doi: 10.1109/ICCT.2010.5688947.

[6] J. F. Yan, L. Pei, Z. W. Chen, and M. J. Wang, "Local adaptive sampling for wireless sensor network powered by energy," *Optik-International Journal for Light and Electron Optics*, vol. 123, no. 23, 2012, pp. 2195—2197. doi: DOI: 10.1016/j.ijleo.2011.11.011.

[7] P. F. Zhang, G. X. Xiao, and H. P. Tan, "Clustering algorithms for maximizing the lifetime of wireless sensor networks with energy-harvesting sensors," *Computer Networks*, vol. 57, no. 14, 2013, pp. 2689-2704.

[8] J. M. L. P. Caldeira, J. A. F. Moutinho, B. Vaidya, P. Lorenz, and J. J. P. C. Rodrigues, "Intra-Body Temperature Monitoring using a Biofeedback Solution," *Second Int. Conf. on eHealth, Telemedicine, and Social Medicine*, St. Maarten, 2010, pp. 119-124. doi: 10.1109/eTELEMED.2010.23.

[9] O. Diallo, J. J. P. C. Rodrigues, M. Sene, and J. W. Niu, "Real-time query processing optimization for cloud-based wireless body area networks," *Information Sciences*, vol. 284, no. 3, 2014, pp. 84-94.

[10] O. R. E. Pereira, J. M. L. P. Caldeira, and J. J. P. C. Rodrigues, "Body sensor network mobile solutions for biofeedback monitoring," *Mobile Networks and Applications*, vol. 16, no. 01, pp. 713-732, 2011. doi: 10.1007/s11036-010-0278-y.

[11] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed diffusion: a scalable and robust communication paradigm for SNs," in *Sixth Annual Int. Conf. on Mobile Comput. and Networking (MobiCOM' 00)*, Boston, MA, USA, pp. 56-67, August 2000.

[12] W. B. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An application specific protocol architecture for wireless microsensor networks," *IEEE Trans. Wireless commu.*, vol. 1, no. 4, pp. 660- 670, Oct. 2002. doi: 10.1109/TWC.2002.804190.

[13] M. J. Handy, M. Haase, and D. Timmermann, "Low energy adaptive clustering hierarchy with deterministic cluster-head selection," in *proc.2002 4th Int. Workshop on Mobile and Wireless Commu. Network*, Stockholm, Sweden, pp. 368-372, 2002.

[14] S. Tyagi and N. Kumar, "A systematic review on clustering and routing techniques based upon LEACH protocol for wireless sensor networks," *Journal of Network and Computer Applications*, vol. 36, no. 2, 2013, pp. 623-645.

[15] O. Diallo, J. J. P. C. Rodrigues, and M. Sene, "Real-time data management on wireless sensor networks: A survey," *Journal of Network and Computer Applications*, vol. 35, 2012, pp. 1013-1021.

[16] S. Jabbar, A. A. Minhas, A. Paul, and S. Rho, "Multilayer cluster designing al-

gorithm for lifetime improvement of wireless sensor networks," *Journal of Supercomputing*, In Press, 2014, doi: 10.1007/s11227-014-1108-y.
[17] V. K. Verma, S. Singh, and N. P. Pathak, "Analysis of scalability for AODV routing protocol in wireless sensor networks," *Optik-International Journal for Light and Electron Optics*, vol. 125, no. 2, January 2014, pp. 748–750.
[18] S. Tyagi, S. K. Gupta, S. Tanwar, and N. Kumar, "EHE-LEACH: Enhanced Heterogeneous LEACH Protocol for Lifetime Enhancement of Wireless SNs," *Int. conf. on advances in compu., commun. and informatics, (ICACCI2013)*, Mysore, India, pp. 1485–1490, 2013.

Manuscript received: 2014-04-21

Biographies

Sudhanshu Tyagi (sudhanshutyagi123@gmail.com) is an associate professor in the Department of Electronics and Communication Engineering, J. P. Institute of Engineering and Technology, India. He received his BEng degree in electronics and telecommunication from North Maharashtra University, India, in 2000. He received his MTech (Hons) degree in electronics and communication engineering from National Institute of Technology, India, in 2005. He is currently pursuing his PhD in electronics and communication engineering from Mewar University, India. He has authored or co-authored six papers in prestigious peer-reviewed journals and conference proceedings. He is on the editorial boards of leading international journals, such as *International Journal of Ad-Hoc and Ubiquitous Computing*, *Inderscience*, and *Journal of the Franklin Institute*. His research interests include lifetime enhancement of homogeneous and heterogeneous WSNs. He is a member of IEEE and IAENG.

Sudeep Tanwar (sudeeptanwar149@gmail.com) is an assistant professor in the Department of Information Technology, Bharat Institute of Technology, India. He received his BTech degree in computer science from Kurukshetra University, India, in 2002. He received his MTech (Hons) degree in information technology from Guru Gobing Singh Inderprastha University, India, in 2009. Since 2010, he has been pursuing his PhD with specialization in wireless sensor network and works under the supervision of Dr Neeraj Kumar of Thapar University, India. He is member of the Computer Society of India, associate member of the Indian Society for Telecommunication Engineers, member of the International Association of Engineers, and a member of the Computer Science Teachers Association, USA. He has authored two papers in peer-reviewed international journal and conference proceedings.

Sumit Kumar Gupta (skgupta123@gmail.com) is currently working as an assistant professor in the Department of Electronics and Communication Engineering, IIMT Engineering College, India. He received his BTech degree in electronic and communication engineering from Chaudhary Charan Singh University, India, in 2002. He is pursuing his MTech degree in electronic and communication engineering from Uttrakhand Technical University, India. He has eight years industry experience in the operations and customer support with Tata Sky, Reliance Big TV, and Videocon D2h. He has authored research papers that have been published in high-quality national and international journals and conferences proceedings.

Neeraj Kumar (neeraj.kumar@thapar.edu) is an associate professor in the School of Computer Science and Engineering, Shri Mata Vaishno Devi University, India. He received his PhD in computer science and engineering from Shri Mata Vaishno Devi University, India, and thereafter worked as a postdoctoral fellow in the UK. He has authored or coauthored more than 70 paper in peer-reviewed journals and conferences proceedings. His research is focused on mobile computing, parallel/distributed computing, multiagent systems, service-oriented computing, and routing and security in wireless adhoc, sensor and mesh networks. He leads the Mobile Computing and Distributed System Research Group. Prior to joining Shri Mata Vaishno Devi University, he worked with HEC Jagadhri and MMEC Mullana India. He has been invited to speak at various IEEE international conferences in India and abroad. He has organized various special sessions in international conferences in his area of expertise in India and abroad. He is a member of the Technical Program Committee of various IEEE-sponsored conferences in India and abroad. He is also on the editorial boards of various international journals. He has been the guest editor of six special issues of international journals. He is senior member of ACM, ACEEE and IACSIT.

Joel J. P. C. Rodrigues (joeljr@ieee.org) is a professor in the Department of Informatics, University of Beira Interior, Portugal. He is also a researcher at the Instituto de Telecomunicações, Portugal. He received his PhD and MSc degrees in informatics engineering from the University of Beira Interior. He received his five-year BSc degree (licentiate) in informatics engineering from the University of Coimbra, Portugal. His main research interests are vehicular communications, sensor networks, eHealth, eLearning, mobile and ubiquitous computing, and cloud technologies. He is the leader of the NetGNA Research Group, chair of the IEEE ComSoc Technical Committee on eHealth, past chair of the IEEE ComSoc Technical Committee on Communications Software, Member Representative of the IEEE Communications Society on the IEEE Biometrics Council, Steering Committee member of the IEEE Life Sciences Technical Community, and officer of the IEEE 1907.1 standard. He is the editor-in-chief of the *International Journal on E-Health and Medical Communications* and *Recent Advances in Communications and Networking Technology*. He is also on the editorial board member of several other journals. He has been the guest editor for a number of journals and has been general chair and TPC chair of many international conferences. He has authored or coauthored more than 350 papers in refereed international journals and conferences proceedings. He has also written a book and holds two patents. He was awarded the Outstanding Leadership Award at IEEE GLOBECOM 2010 as the CSSMA Symposium co-chair and has won several other best-paper awards. He is a licensed professional engineer, a member of the Internet Society, an IARIA fellow, and a senior member of ACM and IEEE.

DOI: 10.3939/j.issn.1673-5188.2014.03.004
http://www.cnki.net/kcms/detail/34.1294.TN.20140829.1054.001.html, published online 29 August, 2014

Prototype for Integrating Internet of Things and Emergency Service in an IP Multimedia Subsystem for Wireless Body Area Networks

Kai-Di Chang[1], **Jiann-Liang Chen**[1], **and Han-Chieh Chao**[2]
(1. Department of Electrical Engineering, National Taiwan University of Science and Technology, Taiwan, China;
2. Department of Computer Science and Information Engineering, Nation ILan University, Taiwan, China)

Abstract

In recent years, the application of the Internet of Things (IoT) has become an emerging business. The most important concept of next-generation network for providing a common global IT platform is combining seamless networks and networked things, objects or sensors. Also, wireless body area networks (WBANs) are becoming mature with the widespread usage of the IoT. In order to support WBAN, the platform, scenario and emergency service are necessary due to the sensors in WBAN being related to wearer´s life. The sensors on the body detect a lot of information about bioinformatics and medical signals, such as heartbeat and blood. Thus, the integration of IoT and network communication in daily life is important. However, there is not only a lack of common fabric for integrating IoT with current Internet and but also no emergency call process in the current network communication environment. To overcome such situations, the prototype of integrating IoT and emergency call process is discussed. A simulated bootstrap platform to provide the discussion of open challenges and solutions for deploying IoT in Internet and the emergency communication system are analyzed by using a service of 3GPP IP multimedia subsystem. Finally, the prototype for supporting WBAN with emergence service is also addressed and the performance results are useful to service providers and network operators that they can estimate their migration to IoT by referring to this experience and experiment results. Furthermore, the queuing model used to achieve the performance of emergency service in IMS and the delay time of the proposed model is analyzed.

Keywords

IoT; WBAN; radio frequency identification (RFID); emergency service; IP multimedia subsystem

1 Introduction

The concept of Internet of Things (IoT) was introduced in 1999 with Auto-ID—the technology that underlies the electronic product code (EPC) system [1]. The EPC system was designed to connect physical objects or devices via radio frequency identification (RFID) and a unique EPC code in the RFID tag. Researchers in Japan also proposed a UID solution that was an early IoT prototype [2]. The IoT concept has expanded very quickly because of new application demands and technological advances [3]–[5]. IoT has been defined in different ways, according to different technologies and points of view. In [6], Thiesse et al. propose solutions based on RFID or EPC. In [7], Broll et al. propose the pervasive service interaction between things. In [8], Vazquez et al. propose a solution that integrates smart objects and mobile services. Most researchers focus on specific applications or functions [9]; however, two areas of commonality are security [10], [11] and network OAM [12]. The Future Internet Assembly was founded by the European Commission to support fundamental and systematic innovation in the future Internet and IoT [13].

Although there has been much research on IoT, the architecture of future IoT has not yet been defined. Objects in IoT have unique identity and virtual personalities operating in smart spaces such as body area network or personal area network through using smart interfaces to connect or communicate with social, cyber and exchange user contexts. IoT technologies can promote the integration of material production with service management, i.e., the integration of physical region, digital region, and cyber region. IoT communication has been mainly supported by the evolution of information processing and services within the ICT industry.

With the popularization of the Internet and mobile communications, IoT come to be regarded as the next wave of IT. The ITU has already stated that machine-to-machine and person-to-machine communications will be extended to a much wider range of devices embedded within the existing Internet. Thus, IoT will be an integral part of any future Internet.

Different types of IoT have already been accepted in many places, but further development of IoT depends greatly on the specification of new technologies, deeper social understanding of IoT-related issues, and stronger legal frameworks for IoT. Standards, reliability, and robustness are all important issues in IoT development, and a standardized architecture is the foundation for all IoT technologies. If there is no definite IoT architecture, it will be difficult to develop and integrate future applications and services.

The rapid development of IoT around the world has largely been led by governments, but industry has also initiated big-budget IoT projects even though the technologies and architecture of IoT are not yet mature. Lack of a clear architecture stymies the development of IoT, so it is very important figure out the future IoT architecture.

IoT is opening up the possibility for new Internet applications, and next-generation networks (NGNs) will have a similar effect. All user requests for multimedia services and applications can be fulfilled by deploying an IP multimedia subsystem (IMS) [14]. IMS applications have grown quickly around the world, and IMS has become a key technology. The main IMS core network is a subsystem bounded with Universal Mobile Telecommunication System (UMTS) to provide end-to-end or core-to-end multimedia services. User equipment is supported so that it can be used in both fixed and mobile networks. There are two types of switching technology in UMTS network: circuit-switching, for voice transmission; and packet-switching, for data transmission. Improvement in the packet-switching domain improves the migration of Internet and mobile communications.

The specifications and standards of IMS are defined and discussed in many international organizations, such as the 3rd generation partner project (3GPP), telecommunications and internet converged services and protocols for advanced networking (TISPAN), ETST and ITU-T for ITU telecommunication standardization sector (ITU-T). The specifications provide a convergence goal for future Internet and service. IMS provides mobility, flexibility and scalability for services and applications. In addition, with the important call session control functions (CSCFs), the IMS core network has the capability of controlling sessions, delivering voice, data, video and message in fixed and mobile network.

Many elements deliver the signaling and control messages to achieve multimedia services and applications. These functions improve the efficiency of IMS. However, with smooth operations in IMS, it is very difficult to avoid the emergency events in daily life. Emergency services are proposed to handle the emergencies. Furthermore, there are many objects for sensing different signal and information in IoT environment that includes the various type of sensor in wireless body are network. The types of sensed message are extensive. Some of the message is highly related to the life of the wearer such as blood, heartbeats. Thus, a reliable message delivery transmission mechanism is necessary for emergency situations. In this research, the IP multimedia subsystem is selected as the fabric for supporting IoT. The emergency service in IMS is very suitable for adapting to WBAN. The emergency signal can be sent through IMS emergency service when the sensors on the wearer sense abnormal bioinformatics and medical signals. In order to analyze the prorogation delay of each element in IMS emergency service, the queuing model is used to achieve the goal and calculate the delay time.

The remainder of this paper is organized as follows. In section 2, we discuss the background of the Internet of Things, IP multimedia subsystem and emergency services. In section 3, we discuss the integrated prototype of IoT and emergency service model. In section 4, we discuss the deployment and performance analysis of the prototype. In section 5, we conclude and discuss future works are illustrated in last section.

2 Background

2.1 Internet of Things

IoT involves many technologies and includes research fields such as network architecture design, sensor and object identification, coding, data transmission, data processing, network planning and link/node discovery, etc. There are four key components of IoT:

1) Sensors with embedded intelligence. A sensor node in a traditional wireless network was designed to sense data, and store and forward the result to a sink. In future IoT, the sensor node will have more intelligent algorithms, cognitive capabilities. Thus, each sensor node will be an intelligent object rather than a simple sensor node.
2) Aggregators, such as hubs, for spokes. The moderate processing concept is included in the aggregators. It's more powerful to handle messages in each place.
3) Ubiquitous network. Objects generate and communicate information about environment or item status when queries triggered. The network connectivity is always on to achieve information communication and data exchanging.
4) Context-aware services. These enhance an object's processing capability and facilitate decision-making between devices without human intervention. Thus, operations are performed automatically.

In addition to the above four key components, there are three important characteristics of IoT:

1) Good cognitive capability and distributed sensing for input/output modules
2) Robust transmission and stable bus for industry communication
3) Smart process and programmable automation controller for adapting to variable data sensing environments.

Ning and Wang [15] propose two IoT architectures: like mankind neural system and social organization framework.

The like mankind neural system comprises M&DC, the brain respond s for objects management and centralized data center; spinal cord, there are distributed control nodes for controlling lowest level sensors; and a network of nerves, deploy IoT network and end-side sensors.

This IoT architecture transmits messages from low-level sensors to mid-level control nodes and top-level M&DC. It receives, translates, and sends back message to sensors to control the things/objects. The M&DC is a centralized data center. It is in charge of processing information, storing data, and its most important task is to manage the IoT network.

In Ning and Wang's design, the social organization framework (SOF) plays three roles in IoT network. For national IoT, the SOF act as national management and data center which is called nM&DC.

For industry IoT, SOF acts as industry management and data center called iM&DC. Finally, for regional IoT, SOF responds to local management and data center called lM&DC. With different types of SOF IoT, each IoT has their different level policies, monitoring, security, and backup of important data. The "Like Mankind Neural System" can be regarded as a single IoT network, then the "Social Organization Framework" comprises many "Like Mankind Neural Systems," such as multi-IoT networks. The main difference between LMNS and SOF is that each IoT network can exchange information with another IoT network. It is like a social network: one LMNS can share sensors with different LMNS. The behavior and status is similar to human conversation in society. However, there is no traffic analysis or model for IoT on internet in their research and they do not mention about their IoT operation management in the two IoT architectures.

2.2 IP Multimedia Subsystem

The detailed specifications of IP multimedia subsystems [16] were proposed by 3GPP. Many multimedia services voice call, data service and can be used smoothly based on the all-IP concept and technology. IMS supports the fixed and mobile devices by using Session Initiation Protocol, and can achieve seamless handover between heterogeneous networks. The concept can be extended to the whole network communication industry. To enable users to use various multimedia services and multimedia content, the construction of NGN infrastructure is very important to operators. Finally, the operator can achieve the management of multimedia service and service control. Zhou and Chao proposed multimedia traffic security architecture [11]. In this study, IMS is also an important architecture for future Internet. Moreover, Chen et al. discuss the importance of IMS in IoT [17]. In addition, the secured IMS environment is also necessary [18] for emergency services.

Session control and resource allocation are managed by the cooperation of SIP servers and other physical network devices. In the IMS layer, the main components, called CSCFs, can process the SIP message in the core network. The serving CSCF (S-CSCF) is the major component of IMS. The S-CSCF is responsible for call session control, handling the registration and authentication control. The proxy CSCF (P-CSCF) is the entry point of users to the IMS, which works as proxy and user agent. It is responsible for forwarding IMS registration message or other IMS requests to IMS core. The interrogation CSCF (I-CSCF) communicates with other network operator, which provides routing query and hides the internal network topology between IMS realms.

Session initiation protocol (SIP) is an open standard proposed by IETF MMUSIC Working Group [19]. SIP provides strong functions for session control in application layer such as session establishment, modification and terminate. The IMS network operators use IMS-SIP as the signaling protocol for providing voice and video multimedia service.

User equipment (UE) can access the IMS through different access networks with the flexibility and scalability of SIP. The signaling time to access IMS is different in different access networks, such as GSM, WLAN and WiMAX. Thus, Wagle et al. compares the IMS access time between GSM and WLAN [16]. The results in [16] show that the GSM-IMS-GSM has longer access time than the WLAN-IMS-WLAN. The delay is affected by different bandwidth and network prorogation delay. Thus, in order to reduce delay, the messages and SIP signaling flow such as call setup, request and reply should be considered. Users can have better quality with lower delay in IMS. Chang et al. proposed an advanced path-migration mechanism [20] for improving signaling efficiency in IMS. However, the research did not take into consideration emergency services.

2.3 Emergency Service

Emergency service in IMS consists of P-CSCF, S-CSCF, emergency-CSCF (E-CSCF), location retrieval function (LRF) and UE. When UE triggers the emergency service, it first calls P-CSCF to decide whether the situation is emergent or not. If the event is truly an emergency, the P-CSCF forwards the event to E-CSCF in the same domain. Then, the E-CSCF can handle the event according to the message content and determine the location of the user from the LRF.

If P-CSCF determines that this is not an emergency, the request is forwarded to the home service proxy; then, the PSAP in home network processes the request. This architecture can handle emergency events effectively. Thus, this service makes the IMS much more reliable and secure to users.

The IMS emergency service in heterogeneous network environments is shown in Chen´s research [21]. It is the combination of IMS core network, Internet and HSDPA data communication by telecom operator, relay node and user. There are three-connection types for users to reach the IMS core: 1) use the 3.5 G data communication service; 2) relay by Bluetooth and cooperate with other user; 3) use wired or Wi-Fi to reach IMS service. Users can use any service in any place at any time with these heterogeneous networks. Also, the emergency service can be launched successfully. Chen et al. also proposed a cooperative IMS network [22] to extend the coverage of emergency service.

3 Integrated Prototype of IoT and Emergency Service Model

IoT is different from traditional wireless sensor networks and current Internet. The new concept should be considered in operations plane and strategy to satisfy the abovementioned four key components and three characteristics of IoT. The architecture and operation management of future internet IoT should be accommodating with IoT related workflow and tasks.

The architecture for network OAM depends on the network scale. For example, the IoT architecture and the work-processing model shall be centralized for small-scale system. Thus, the consumed resource can be reduced. In order to provide enough system capacity, the distributed model needs to be considered if IoT is deployed in a large-scale network such as multiple applications environment or ubiquitous system.

Unlike current Internet and mobile communications, IoT architecture for operation management should be flexible, compromising and ubiquitous. Thus, each country or their industry can easily access a suitable IoT architecture or exclusive IoT architecture and communicate with other IoT networks.

3.1 IoT Operations

In order to achieve IoT operations management, the basic concept of IoT system architecture should be determined first. The layered architecture of IoT can be extended from WSN, and the extended architecture can be separated into three layers.

Application layer-operators use the collected information to deploy their service and applications, such as intelligent traffic management, smart home [23], [24], smart gird, long-distance healthcare [25] environmental protection, mining monitoring, remote nursing, safety defense, and smart government.

Network Layer-there are data center, core network, network backbone, mobile phone networks, fixed telephone networks, broadcasting networks, and closed IP data networks for each carrier in this layer. All traffic is IP based and will be routed to the suitable destination in this layer.

Sensing Layer-the operators deploy RFID related devices such as sensor, tag, sensor gateway, smart terminal, and IoT gateway. Those objects in this layer handle the data cognition and information collection.

Thus, the operation management would focus on the functions in each layer.

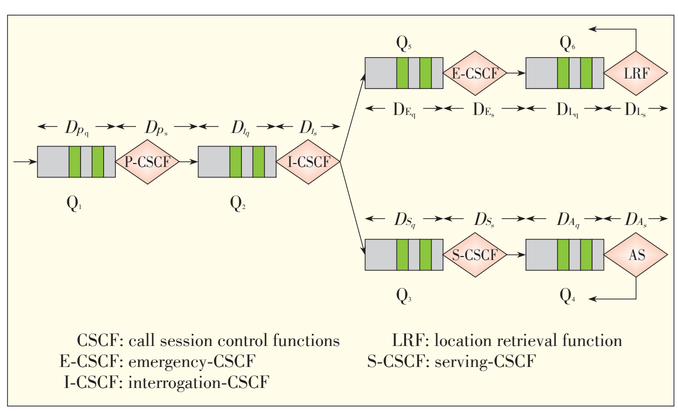
3.2 IoT Business Operation Support Platform

Although there is no definitely specification for IoT, Qian et al. propose a Business Operation Support Platform (BOSP) for IoT management in business operation [26]. The general architecture of BOSP is shown in Qian's design:

1) data interface. This interface receives IP packets over interconnected heterogeneous networks for all IP-based networks, such as broadcasting networks, fixed networks and mobile/wireless networks. It then forward packets to corresponding destinations.
2) application interface. This interface forwards data to application servers and provides specific functions to facilitate application systems.
3) existing abilities entrance. This entrance makes carrier's existing abilities work directly. For example, the voice call service, video call, short messaging service, multimedia messaging service, location based service, conference service and some multimedia service can be used through existing abilities entrance.
4) existing BSS/OSS interface. This interface is used to visit billing/charging system such as CRM (customer relationship management) system, network management system, operational data store and operation analysis system through carrier's ESB (enterprise service bus) to achieve accounting functions. Thus, the users used service and resource cost can be billed for business operation.

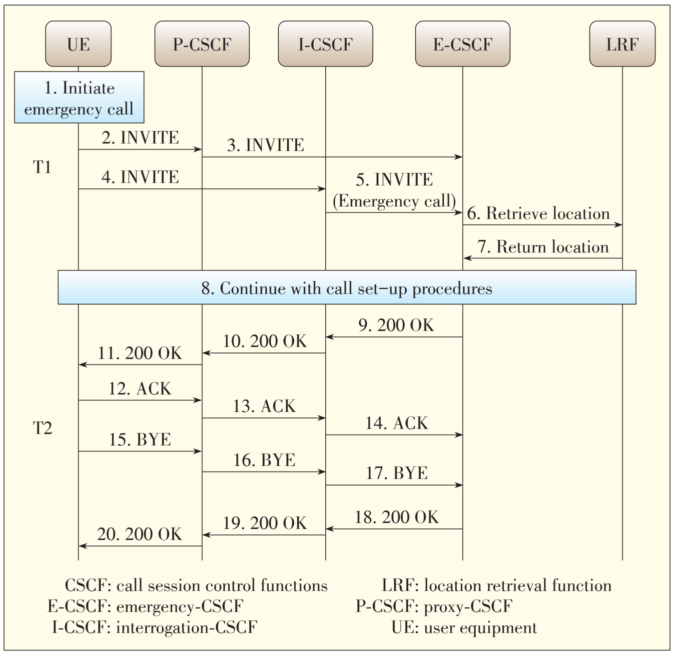
3.3 Queuing Time in Emergency Service

The emergency service mainly consists of elements in IMS. All the elements have different roles, such as P-CSCF, I-CSCF, S-CSCF and E-CSCF. The Poisson distribution is used to calculate the probability of service transfer and queue transaction. Then, the delay cost in each queue is evaluated using a queuing model. **Fig. 1** shows the queuing model for emergency service. This is an M/M/1/∞/∞ model. In the beginning, UEs have to wait in Q_1. Then, they communicate with P-CSCF in order to connect to IMS. If the UE is in the front of the queue, it cannot enter the service queue. The P-CSCF first determines wither this is an emergency. If it is an emergency, the request is forwarded to Q_5 through Q_2 (I-CSCF). If not, the request is forwarded to Q_3 through Q_2 (S-CSCF) to be a normal service. **Fig. 2** shows the delay in different components. The delay



▲**Figure 1. Queuing model for emergency service.**

Kai-Di Chang, Jiann-Liang Chen, and Han-Chieh Chao



▲**Figure 2. Emergency service call setup procedure.**

mainly consists of D_{Pq}, D_{Ps}, D_{Iq}, D_{Is}, D_{Eq}, D_{Es}, D_{Lq}, D_{Ls}, D_{Sq}, D_{Ss}, D_{Aq} and D_{As}, the detailed explanations of notations are shown in **table 1**.

All the requests are analyzed by using the standard M/M/1∞/∞ queuing model. The arrival rates in the IMS core network are denoted λ_{PP}, λ_{II}, λ_{SS}, λ_{AA}, λ_{EE}, λ_{LL}。 The service rates of elements in the IMS core network are denoted μ_{PP}, μ_{II}, μ_{SS}, μ_{AA}, μ_{EE}, μ_{LL}. All the requests before entering the element have to wait until the element (P-CSCF, I-CSCF, S-CSCF, AS, E-CSCF, and LRF) is free to handle the request.

$$D_{P_q} = \frac{\lambda_{PP}}{\mu_{PP}(\mu_{PP} - \lambda_{PP})} \tag{1}$$

$$D_{I_q} = \frac{\lambda_{II}}{\mu_{II}(\mu_{II} - \lambda_{II})} \tag{2}$$

$$D_{S_q} = \frac{\lambda_{ss}}{\mu_{ss}(\mu_{ss} - \lambda_{ss})} \tag{3}$$

$$D_{A_q} = \frac{\lambda_{AA}}{\mu_{AA}(\mu_{AA} - \lambda_{AA})} \tag{4}$$

$$D_{E_q} = \frac{\lambda_{EE}}{\mu_{EE}(\mu_{EE} - \lambda_{EE})} \tag{5}$$

$$D_{L_q} = \frac{\lambda_{LL}}{\mu_{LL}(\mu_{LL} - \lambda_{LL})} \tag{6}$$

After establishing the link, the request is processed by P-CSCF, I-CSCF, S-CSCF, AS, E-CSCF, LRF. From (7)–(12), the process time in each element can be achieved.

$$D_{P_s} = \frac{1}{\mu_{PP}(\mu_{PP} - \lambda_{PP})} \tag{7}$$

$$D_{I_s} = \frac{1}{\mu_{II}(\mu_{II} - \lambda_{II})} \tag{8}$$

$$D_{S_s} = \frac{1}{\mu_{SS}(\mu_{SS} - \lambda_{SS})} \tag{9}$$

$$D_{A_s} = \frac{1}{\mu_{AA}(\mu_{AA} - \lambda_{AA})} \tag{10}$$

$$D_{E_s} = \frac{1}{\mu_{EE}(\mu_{EE} - \lambda_{EE})} \tag{11}$$

$$D_{L_s} = \frac{1}{\mu_{LL}(\mu_{LL} - \lambda_{LL})} \tag{12}$$

Before entering the P-CSCF, I-CSCF, S-CSCF, AS, E-CSCF and LRF, the average waiting requests can be obtained according to (13)–(18).

$$L_{PP_q} = \lambda_{PP} \times D_{P_q} \tag{13}$$

$$L_{II_q} = \lambda_{II} \times D_{I_q} \tag{14}$$

$$L_{SS_q} = \lambda_{SS} \times D_{I_q} \tag{15}$$

$$L_{AA_q} = \lambda_{AA} \times D_{A_q} \tag{16}$$

$$L_{EE_q} = \lambda_{EE} \times D_{E_q} \tag{17}$$

$$L_{LL_q} = \lambda_{LL} \times D_{L_q} \tag{18}$$

The average waiting number in P-CSCF, I-CSCF, S-CSCF, AS, E-CSCF and LRF is given by (19)–(24)

$$L_{PP_s} = \lambda_{PP} \times D_{P_s} \tag{19}$$

$$L_{II_s} = \lambda_{II} \times D_{I_s} \tag{20}$$

$$L_{SS_s} = \lambda_{SS} \times D_{I_s} \tag{21}$$

$$L_{AA_s} = \lambda_{AA} \times D_{A_s} \tag{22}$$

$$L_{EE_s} = \lambda_{EE} \times D_{E_s} \tag{23}$$

$$L_{LL_s} = \lambda_{LL} \times D_{L_s} \tag{24}$$

The overall delay can be calculated by the count waiting time of system in the queue and service time. Thus, D_{T1} is the total delay in the normal situation. D_{T2} is the total delay in an emergency situation. The total delay of IMS is given by (25)

▼**Table 1. Delay notations**

Notation	Delay
D_{Pq}	Queuing time before entering the system.
D_{Ps}	Queuing time before processing/serving by P-CSCF.
D_{Iq}	Queuing time in Queue 2.
D_{Is}	Queuing time for service after entering I-CSCF.
D_{Sq}	Queuing time in Queue 3 before entering E-CSCF.
D_{Ss}	Queuing time for processing/service by S-CSCF.
D_{Aq}	Queuing time before entering system in Queue 4.
D_{As}	Queuing time before processing/serving by application server.
D_{Eq}	Queuing time in Queue 5 to emergency system.
D_{Es}	Queuing time for service time after entering E-CSCF system.
D_{Lq}	Queuing time in Queue 6.
D_{Ls}	Queuing time when entering LRF to get the location of user.

CSCF: call session control functions
E-CSCF: emergency-CSCF
I-CSCF: interrogation-CSCF
P-CSCF: proxy-CSCF
S-CSCF: serving-CSCF

and (26):

$$D_{T1}=D_{Pq}+D_{Ps}+D_{Iq}+D_{Is}+D_{Sq}+D_{Ss}+D_{Aq}+D_{As} \quad (25)$$
$$D_{T2}=D_{Pq}+D_{Ps}+D_{Iq}+D_{Is}+D_{Eq}+D_{Es}+D_{Lq}+D_{Ls} \quad (26)$$

In order to achieve the arrival rate in different queue, the different arrival rate is considered in **Table 2** and **Table 3**. Table 2 shows the message arrival rate with Poisson distribution when λ = 50. Table 3 shows the service rate of system elements with Poisson distribution when λ = 50.

The arrival rate λ and service rate μ in Table 2 and Table 3 are measured with Wireshark. These values can be used to estimate the delay time of system through the (1)–(26):

$$D_{Pq}=\frac{12\lambda_{PP}}{\mu_{PP}(\mu_{PP}-\lambda_{PP})}+\frac{2\times 0.78\times\lambda_{PP}}{\mu_{PP}(\mu_{PP}-0.78\times\lambda_{PP})}$$
$$D_{Ps}=\frac{2}{\mu_{PP}-\lambda_{PP}}+\frac{2}{\mu_{PP}-0.78\times\lambda_{PP}}=5.7\text{ ms}$$

$$D_{Iq}=0.97\text{ ms};\ D_{Is}=9.8\text{ ms}$$
$$D_{Eq}=228.5\text{ ms};\ D_{Es}=295.2\text{ ms}$$
$$D_{Lq}=0.32\text{ ms};\ D_{Is}=0.55\text{ ms}$$

The total delay in an emergency system is D_{T2} = 546.3 ms.

In this section, IoT Operations, IoT Business Operation Support Platform and Queuing time in Emergency Service are discussed.

4 Deployment and Performance Analysis

The proposed prototype system is shown in **Fig. 3**, which is consists of WBAN, LAN, WAN, IoT domain and IMS domain. Then, the deployment evaluation, traffic analysis and emergency signaling analysis are discussed.

4.1 IoT Deployment Evaluation

In the current environment, the main consideration of deploying IoT network is connecting low-layer objects to the Internet. Also, there are lack of specifications and standards for IoT. The business operation model or platform is not shown in the current Internet. In order to evaluate IoT deployment, we construct a bootstrap platform and map it with Ning and Wang's Like Mankind Neural System [15] and combined with wireless sensor body network, the OPNet network modeler is taken by this research. Finally, the environment for integrating IoT and IMS for WBSN are proposed as an environment.

4.2 IoT Traffic Analysis

In this simulation, each object sends a 1024 bit packet per second, which stores sensed information, to the coordinator in this IoT bootstrap platform. Then the coordinator handles these messages and feeds back to each object. We measure number of hops, traffic to coordinator, router and object, and finally the average end-to-end delay.

The number of hops for data transmission is shown in **Fig. 4**. In this scenario, most objects send their data to the coordinator via one hop. However, some data is delivered via more than two hops. The reason for this is the native limitation of IoT router and the distance from object to coordinator. This causes the information to be delivered through other objects.

The traffic in bootstrap platform is drawn in **Fig. 5**. The IoT

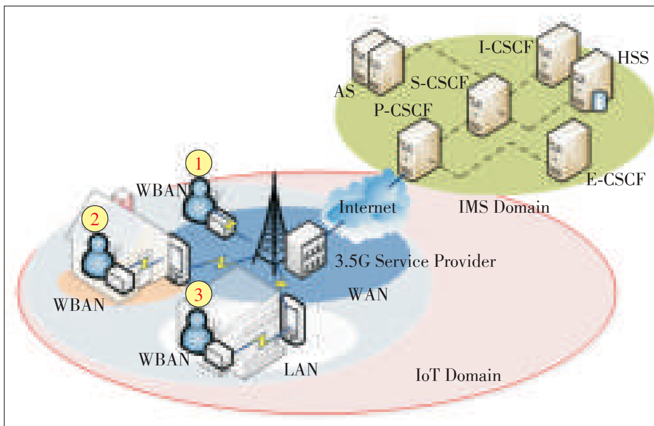
▼**Table 2. The inter arrival rate of different message (λ = 50)**

Delay	INVITE	100	101	200
Q_1	$\lambda_{PP}=\lambda$	$\lambda_{PP}=\lambda$	$\lambda_{SS}=0.78\lambda$	$\lambda_{SS}=0.78\lambda$
Q_2	$\lambda_{II}=\lambda$	$\lambda_{II}=\lambda$	$\lambda_{SS}=0.78\lambda$	$\lambda_{SS}=0.78\lambda$
Q_3	$\lambda_{SS}=\lambda$	$\lambda_{SS}=0.78\lambda$	$\lambda_{SS}=0.78\lambda$	$\lambda_{SS}=0.78\lambda$
Q_4	$\lambda_{AA}=0.96\lambda$			

▼**Table 3. The service rate of each function (λ = 50)**

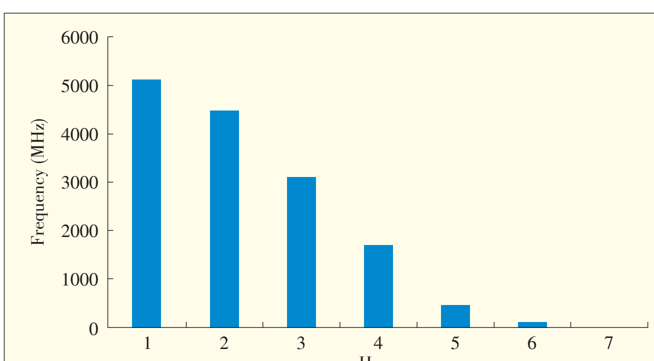
P-CSCF	I-CSCF	S-CSCF	E-CSCF	AS	LRF
$\mu_{PP}=732$	$\mu_{II}=450$	$\mu_{SS}=56$	$\mu_{EE}=60$	$\mu_{AA}=41$	$\mu_{LL}=754$

CSCF: call session control functions
E-CSCF: emergency-CSCF
I-CSCF: interrogation-CSCF
LRF: location retrieval function
P-CSCF: proxy-CSCF
S-CSCF: serving-CSCF



CSCF: call session control functions
E-CSCF: emergency-CSCF
HSS: home subscriber server
I-CSCF: interrogation-CSCF
IMS: IP multimedia subsystem
P-CSCF: proxy-CSCF
S-CSCF: serving-CSCF
WBAN: wireless body area networks

▲**Figure 3. Overview of proposed prototype system.**



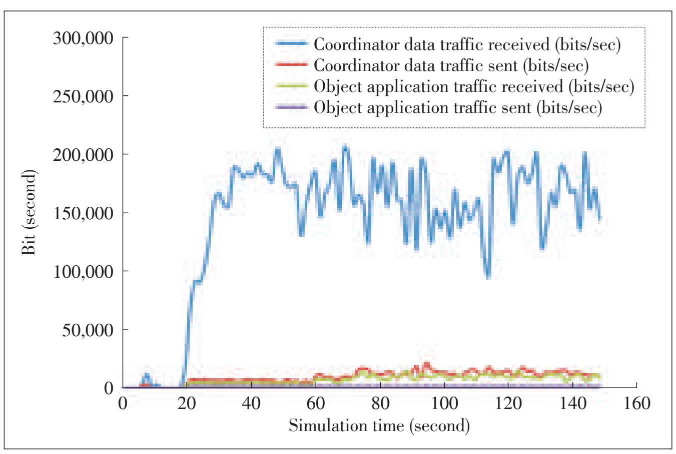
▲**Figure 4. Data transmission hops.**

coordinator receives most data from objects. The IoT router forwards data from objects to the coordinator and vice versa.

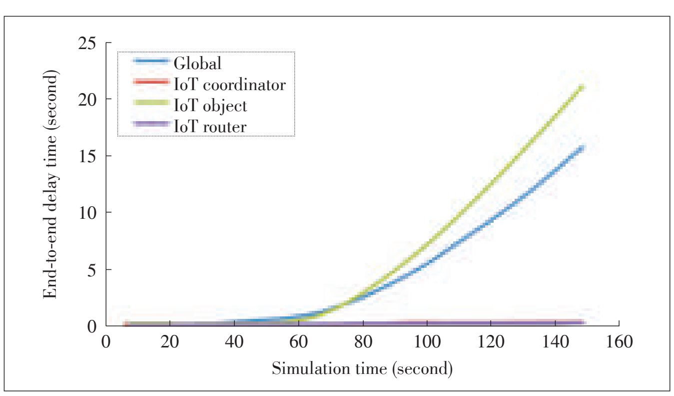
The average end-to-end delay is given in **Fig. 6**. In the beginning, the delay is quite small; however, traffic increases and delay rises violently. The reason is that the queue length in each router and coordinator is fixed. Finally, the delay increases with the continually incoming traffic.

4.3 Emergency Signaling Analysis

In this paper, the call setup time of emergency service is discussed and analyzed. The main signaling of emergency call is shown in Fig. 2 [27]. It includes three phases that can be separated by T_1 and T_2. T_1 begins from UE dials emergency number such as 911 to lunch emergency service. The P-CSCF is responsible for delivering messages from UE to IMS. This confirms that the request is an emergency. If the location of an emergency event was not clear, the emergency CSCF triggers the LRF to obtain location of the UE. Then, the emergency service starts with the detailed event, message, and location. Thus, the different phase of emergency service is obtained from the following definition. T_1 is the time from INVITE message to call setup procedures. T_2 is the time for the call setup procedure to be completed. Finally, the total delay time of an emergency call can be calculated from the relationship between T_1 and T_2: $D = T_2 - T_1$.



▲**Figure 5. Traffic in components' MAC layer.**

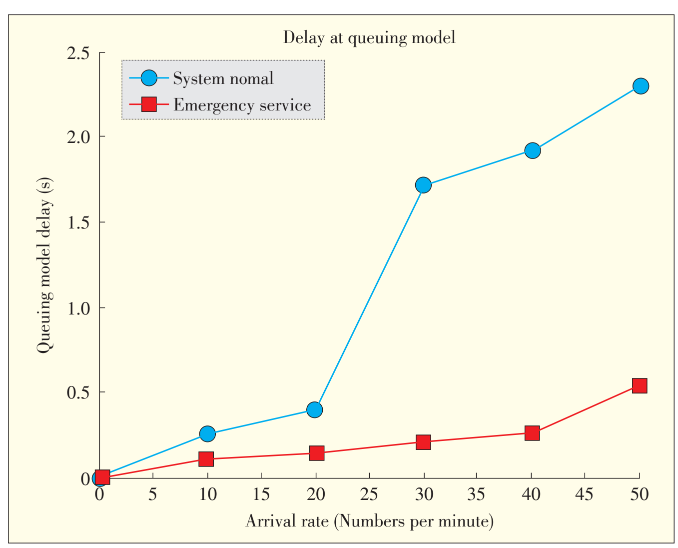


▲**Figure 6. Average end-to-end delay in each component.**

In an emergency, users make an emergency call in a short time, and the request needs to be made as soon as possible. The result of emergency service and proposed queuing model is shown in Fig. 4. From the simulation result in **Fig. 7**, the call setup time in normal system and emergency service are compared. Then, the call arrival rate is increased to 10, 20, 30, 40 and 50. The results show the call setup delay in a normal system increases when the arrival rate is higher than 30. All the call setup time in emergency service is lower than the normal system. Thus, the emergency can provide lower delay and process the emergency call in a short time.

5 Conclusion

In this paper, the prototype of IoT bootstrap platform for supporting WBAN in the network modeler is evaluated. The current Internet is insufficient to maintain the operation quality when constructing and deploying the IoT networks from traffic analysis results. In our opinion, 3GPP IP multimedia subsystem can be a fabric and a suitable core network for integrating IoT mobile communication network, Internet [17]. Furthermore, the emergency service in IMS is discussed. The detailed signaling and application of emergency elements are analyzed with queuing model. Then the network scenario and element are considered to calculate the call setup delay and compare the simulation result in both a normal IMS system and IMS with emergency service. From the results, the call setup time with different arrival rate in emergency service is always lower than normal system. Thus, the emergency service can be used to establish emergency calls, deliver emergency messages, and make it useful in emergencies. The goal of this research is to reduce the call setup time and improve the performance of the emergency service. Finally, for obtaining better, more suitable



▲**Figure 7. The call setup delay time in both emergency service and normal system.**

support for WBAN, the tighter integration of IoT and emergency service through 3GPP IMS would be evaluated in our future work.

Acknowledgements

This research was partly funded by Ministry of Science and Technology of R.O.C. under grants no. NSC 101-2221-E-197-008-MY3.

Reference

[1] *The EPCglobal Architecture Framework (final version 1.3, 2009)*, EPCglobal, 2009.
[2] Huansheng Ning, "Unit and Ubiquitous Internet of Things," CRC Press, April 4, Florida, USA, 2013.
[3] J. P. Conti, "The Internet of things," *Commun. Engineer*, vol. 4, pp. 20–25, 2006.
[4] ITU, "The Internet of things," ITU International Reports, 2005.
[5] A. Bassi and G. Horn, "Internet of Things in 2020 the roadmap in the future," INFSO D.4 Networked Enterprise&RFID INFSO G.2 Micro&Nanosystems(DG INFSO), in co-operation with the RFID working group of the european technology platform on smart systems intergration (EPoSS), version 1.1-27 May, 2008.
[6] F. Thiesse, C. Floerkemeier, M. Harrison, F. Michahelles, and C. Roduner, "Technology, standards, and real-world deployments of the EPC network," *IEEE Internet Computing*, vol. 13, pp. 36–43, 2009. doi: 10.1109/MIC.2009.46.
[7] G. Broll, E. Rukzio, M. Paolucci, M. Wagner, A. Schmidt, and H. Huβmann, "Perci: pervasive serviceinteraction with the Internet of things," *IEEE Internet Computing*, vol. 13, pp. 74–81, Nov.–Dec. 2009. doi: 10.1109/MIC.2009.120.
[8] J. I. Vazquez, J. R. D. Garibay, X. Eguiluz, I. Doamo, S. Renteria, and A. Ayerbe, "Communication architectures and experiences for web-connected physical smart objects," in *Proc. 2010 IEEE International Conference on Pervasive Computing and Communications Workshops*, Mannheim, Germany, 2010, pp. 684–689.
[9] L. Yan, Y. Zhang, Laurence, T. Yang, and H. Ning, *Internet of Things: From RFID to the Next-Generation Pervasive Networked Systems*. British: Auerbach Publications, 2008.
[10] E. Renault, "Toward a security model for the future network of information," in *Proc. 4th International Conference on Ubiquitous Information Technologies & Applications*, Fukuoka, Japan, Dec. 2009, pp. 1–6. doi: 10.1109/ICUT.2009.5405676.
[11] L. Zhou and H. C. Chao, "Multimedia traffic security architecture for the internet of things," *IEEE Network*, vol. 25, no. 3 pp. 35–40, 2011. doi: 10.1109/MNET.2011.5772059.
[12] H. Ning, N. Ning, S. Qu, Y. Zhang, and H. Yang, "Layered structure and management in Internet of things," in *Proc. Future Generation Communication and Networking*, Jeju-Island, Korea, Dec. 2007, pp. 386–389.
[13] G. Tselentis, A. Galis, A. Gavras, S. Krco, V. Lotz, E. Simperl, Burkhard Stiller, and T. Zahariadis, *Towards the Future Internet-Emerging Trends from European Research*. IOS Press, 2010.
[14] *3rd Generation Partnership Project; Technical Specification Group Services and Systems Aspects; IP Multimedia subsystem (IMS); Stage 2*, Technical Specification 3G TS 23.228 version 6.10.0 (2005-08), 2005.
[15] H. Ning and Z. Wang, "Future internet of things architecture: like mankind neural system or social organization framework?" *IEEE Communications Letters*, vol. 15, pp. 461–463, 2011. doi:10.1109/LCOMM.2011.022411.110120.
[16] Shyam S. Wagle, Nikolaos Doulamis, Minesh Ade, and M. Ghazanfar Ullah, "Performance analysis of IP multimedia subsystem (IMS) signaling in heterogeneous access networks," *2011 2nd International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology (Wireless VITAE)*, Chennai, India, Feb. 28-Mar. 3, 2011.
[17] C. Y. Chen, H. C. Chao, T. Y. Wu, C. I Fan, J. -L. Chen, Y. -S. Chen, and J. -M. Hsu, "IoT-IMS communication platform for future internet," *International Journal of Adaptive, Resilient and Autonomic Systems (IJARAS)*, IGI Global, vol. 2, no. 4, pp. 74–94, 2011. doi: 10.4018/jaras.2011100105.
[18] C. Y. Chen, K. D. Chang, and H.C. Chao, "Transaction pattern based anomaly detection algorithm for IP multimedia subsystem," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 1, pp. 152–161, March 2011. doi: 10.1109/TIFS.2010.2095845.
[19] *SIP: Session Initiation Protocal*, IETF RFC 3261, June 2002.
[20] K. D. Chang, C. Y. Chen, S.W. Hsu, H. C. Chao, and J. L. Chen, "Advanced path-migration mechanism for enhancing signaling efficiency in IP multimedia subsystem," *KSII Transactions on Internet and Information Systems*, vol. 6, no. 1, pp. 305–321, January 2012.
[21] C. Y. Chen, K. D. Chang, H.C. Chao, and S.Y. Kuo, "Ubiquitous IMS emergency services over cooperative heterogeneous networks," in *Proc. 2009 ACM International Wireless Communications and Mobile Computing Conference, IWCMC 2009*, Leipzig, Germany, June 21–24, 2009, pp. 968–972. doi: 10.1145/1582379.1582591.
[22] C. Y. Chen, K. D. Chang, H. C. Chao, and J. L. Chen, "Extending emergency services coverage in cooperative IMS networks," *International Journal of Autonomous and Adaptive Communications Systems (IJAACS)*, vol. 5, no. 4, pp. 417–434, 2012.
[23] Y. W. Ma, C. F. Lai, C. C. Hu, M. C. Chen, and Y. M. Huang, "RFID-based seamless multimedia services for smart homes," *International Journal of Internet Protocol Technology*, vol. 4, no. 4, pp. 232–239, 2009.
[24] Y. Wang, and K. Li, "Topology mining of sensor networks for smart home environments," *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 7, no. 3, pp. 163–173, 2011.
[25] Y. M. Huang, M. Y. Hsieh, H. C. Chao, S. H. Hung, and J. H. Park, "Pervasive, secure access to a hierarchical sensor-based healthcare monitoring architecture in wireless heterogeneous networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, pp. 400–411, 2009. doi: 10.1109/JSAC.2009.090505.
[26] X. C. Qian and J. D. Zhang, "Study on the structure of 'Internet of Things (IOT)' business operation support platform," in *Proceedings of 2010 12th IEEE International Conference on Communication Technology (ICCT)*, Nanjing, China, Nov. 2010, pp. 1068–1071.
[27] *3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Feasibility Study on Voice Call Continuity Support for Emergency Calls (Release 9)*, Technical Report 3G TR 23.826 version 9.0.0 (2009 - 03), 2009.

Manuscript received: 2014-04-03

Biographies

Kai-Di Chang (kedy@ieee.org) received his BS degree in electrical engineering from National Dong Hwa University, Taiwan, in 2007. He received his MS degree from the Institute of C0omputer Science and Information Engineering, National I-Lan University, Taiwan. He is currently pursuing his PhD degree in electrical engineering at National Taiwan University of Science and Technology. He is a student member of IEEE. His research interests include cloud computing, IP multimedia subsystems, Internet of Things, and network security. He is also a researcher at United Daily News Digital Taiwan, where his primary role is to research and develop ICT technologies according to his research interests. His professional certifications include IRCA Information Security Management Systems (ISMS) Auditor / Lead Auditor (ISO 27001:2005), 2009; Cisco Certified Academy Instructor (CCAI) for CCNA Security (CCNAS), 2010; and BS 10012 Personal Information Management System Lead Auditor (TLATPI0109), 2013.

Jiann-Liang Chen (lchen@mail.ntust.edu.tw) received his PhD degree in electrical engineering from National Taiwan University, in 1989. Since August 1997, he has worked in the Department of Computer Science and Information Engineering, National Dong Hwa University. He is currently a professor and vice dean of Science and Engineering College at that university. He is also a full professor in the Department of Electrical Engineering, National Taiwan University of Science and Technology. His current research interests include cellular mobility management, digital home network, telematics applications, cloud computing, and RFID middleware design. He is an IEEE Senior Member and UK BCS fellow. He has published more than 150 papers in journals and conferences proceedings and also holds several patents.

Han-Chieh Chao (hcc@niu.edu.tw) received his MS and PhD degrees in Electrical Engineering from Purdue University in 1989 and 1993. He is a jointly appointed professor in the Department of Electronic Engineering and Institute of Computer Science and Information Engineering, National Ilan University, Taiwan. He also holds a joint professorship in the Department of Electrical Engineering, National Dong Hwa University, Taiwan. His research interests include high-speed networks, wireless networks, and IPv6 based networks and applications. He is an IEEE senior member and fellow of the IET and BCS.

DOI: 10.3969/j. issn. 1673-5188. 2014. 03. 005
http://www.cnki.net/kcms/detail/34.1294.TN.20140903.1443.002.html, published online 3 September, 2014

Smart Body Sensor Object Networking

Bhumip Khasnabish
(Strategy Planning and Standards Development, ZTE (TX) Inc., USA)

Abstract

This paper discusses smart body sensor objects (BSOs), including their networking and internetworking. Smartness can be incorporated into BSOs by embedding virtualization, predictive analytics, and proactive computing and communications capabilities. A few use cases including the relevant privacy and protocol requirements are also presented. General usage and deployment etiquette along with the relevant regulatory implications are then discussed.

Keywords

body sensor objects; body sensor networking object; virtualization; predictive analytics; body sensor usage etiquette

1 Introduction

This paper discusses the features, functions, and feasible implementation of smart body sensor objects (BSOs). Localized networking and relevant internetworking (using IP as glue) are also described for mobility and enterprise services [1]. In [2]–[5], the major constraints on implementing smart objects were memory, storage, bandwidth, code size, footprint (both size and weight), and energy consumption and conservation. We propose incorporating smartness into wearable BSOs by embedding virtualization, predictive analytics, and proactive computing and communication. Virtualization [6]–[10] of memory, storage and networking is prevalent in today's computing and communication environments, and it is only natural to extend virtualization into the emerging field of smart objects and networks. Prediction, such as Bayesian inference and associated analytics [1]– [10], is used to pre-position computing and communication facilities in addition to better managing network resources. Finally, proactive computing and communication enables efficient object provisioning and configuration management so that smart objects are deployed as smoothly as possible. A recent study revealed that the worldwide wearable computing market is growing, with almost 20 million shipments expected in 2014 [11].

The rest of this paper is organized as follows. In section 2, we present a few use cases for BSOs. In section 2, we present a general BSO architecture. In section 4, we describe some implementations of BSOs. In section 5, we discuss smart BSO network (SBSON) models. In section 6, we discuss monitoring models based on body sensor network objects. In section 7, we give the results of experiments conducted on a BSO network. In section 8, we conclude the paper and discuss future work directions.

2 Use Cases

2.1 Lifestyle Monitoring

Smart body sensors can be used to monitor lifestyle activities, such as walking or jogging, yoga, swimming, talking, eating, or other conventional or non-conventional body movements. It can also be used to monitor the triggers of allergic reactions. Such monitoring should be followed by passive reporting of the results (raw or correlated, in audio or text format) to pre-selected devices, which may include a device that a device worn by the person themselves. Sensors used for lifestyle monitoring should be compact, comfortable, and non-intrusive.

2.2 Vital–Organ Monitoring

Smart body sensors can be used to monitor events and triggers related to vital organs. Such events and triggers can be monitored and reported at a regular (default) rate or a higher-than-normal rate. Triggers may be caused by abnormal activity in one organ, and this may affect the functioning of other organs. It is possible to develop templates for monitoring vital organs and events, and these templates can be used across different autonomous systems for rapid diagnosis and corrective measures. As with lifestyle-monitoring sensors, monitoring should be followed by active or passive reporting of the raw or correlated results. The objective is to stop the person's vital organ from deteriorating while ambulance officers are on their way. It may be desirable that only sensors non-intrusive to others' personal space be used for vital organ monitoring. If by mistake the sensors record any non-personal data or trigger, these should probably be treated as the property of a third party.

2.3 Black Box

A black box may be for private use (Pr.BB) or public use (Pu.BB). A Pr.BB can be used for smart-sensor monitoring and recording of heart and pulse rates, blood pressure, blood sugar levels, limb movements, speech, or even taking pictures of people you are talking to. A Pu.BB can be used for smart-sensor monitoring of what you are wearing, the street you are in, where you have come from, who you have met, and what you talked about and ate. This may also include watching the watcher, monitoring the potential perpetrator, analyzing and correlating events for taking proactive actions (passive or active step for protection), etc. There is a fine line between what is considered private and what is considered public, and laws and regulations may not be very clear in this area just yet. Further research is needed to provide local and universal guidelines in this area.

3 Body Sensor Object

In general, a BSO comprises computing, communication, and service-specific sensors and application modules. Both computing and communication modules have generic and specialized submodules. These submodules are special-purpose components and entities related to the services provided by the sensor. The submodules may be related to the sensing, monitoring, and analysis of temperature, images, graphics, pressure, vibration, or pollution.

The base Apps module of the BSO is concerned with data/information collection, correlation, and decision-making (e.g., comparison with a pre-set array of thresholds). This can be augmented by advanced or application-specific modules for generating actionable, outlier-free intelligent data and alarms for local or remote use. Analytical and prediction capability can be extended by using virtualized computing and processing modules according to the availability of resources. These capabilities reflect the smartness of the BSO. Appropriate lightweight application programming interfaces (APIs), such as HTTP publish/subscribe model, are used for locating, accessing, reserving, and using virtual communication port, logical channel, memory, processing, and storage.

The concept of BSO can be easily extended to cover body sensor network objects (BSNOs). The essential component of a BSNO is the networking module and component of the object. Bandwidth and security of the network channel are of paramount importance in this scenario. As in [4], low-overhead processing and energy-efficient authorization, authentication, and encryption mechanisms must be used in constrained environments such as a body sensor network.

Fig. 1 shows high-level functional architecture of a BSNO. **Fig. 2** shows the modules for embedding smartness in a BSNO. **Fig. 3** and **Fig. 4a** shows the virtualization of the entities of a BSO and BSNO through appropriate APIs.

Fig. 4b shows personal data store (PDS) with collection, categorization, storage, and APIs for providing appropriate access. Data can be collected from both private and public interactions of a person with applications and services, e.g., email, web access and browsing; and system, e.g., census and blogs. Maintenance, which includes archiving and categorization, can be based on different criteria. Access to the PDS may be for black box and other applications, and different APIs can be used after appropriate embedded or on-demand authentication.. Although further granularization is possible, at present, personal data can be categorized as private, public, secret, or top-secret (Fig. 4b).

4 Body Sensor Implementation

A basic BSO contains the same computing and communica-

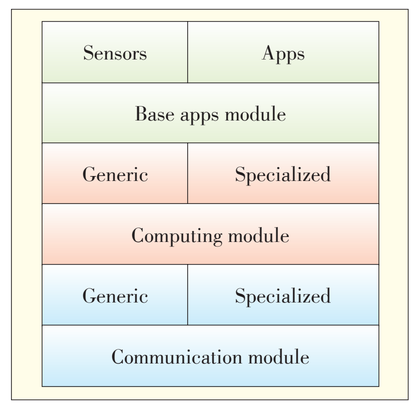
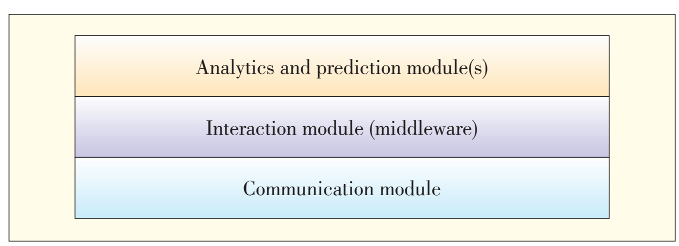
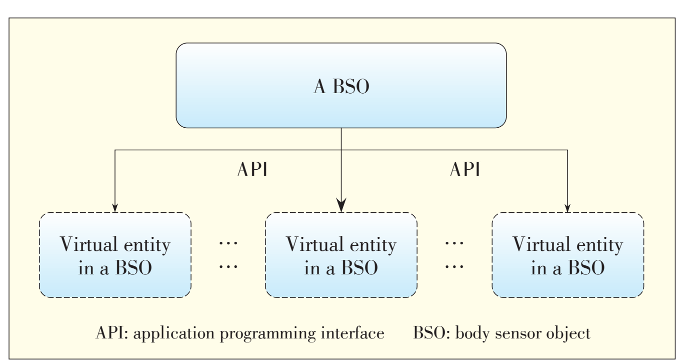


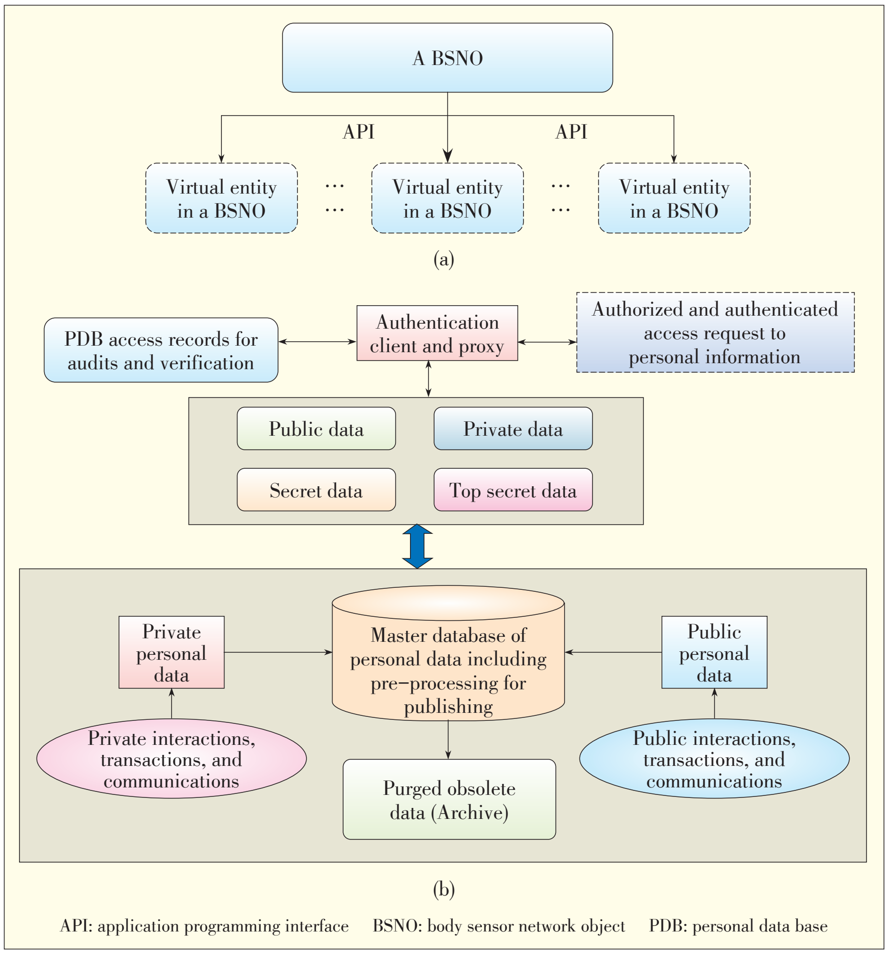
Figure 1. ▶ High-level functional architecture of a smart body sensor network object.



▲Figure 2. Supplimentary body sensor network object modules for embedding smartness.



▲Figure 3. Virtualized entities in a body sensor object.



▲**Figure 4. a)Virtualized entities in a body sensor network object, and b) architecture for personal data store.**

tion elements, but the sensing entities may depend on the targeted purpose. A radio frequency identifier (RFID) tag is also helpful and can be embedded into each of the body sensors.

When the basic elements are present, the rest of the entities can be implemented in different shapes and forms to fit the wearable device. For example, the shape of a wearable electrocardiogram (ECG) sensor can be adjusted according to chest measurements (**Fig. 5**). Similarly, a wearable pulse meter may contain a wristband and device that fits onto the finger tips (**Fig. 6**). **Fig. 7** and **Fig. 8** shows Google Glass smart goggles that provide interfaces for recording and take snapshots of what one is seeing. It can also display stored, live, and on-demand video because antennas, speakers, and microphones are embedded in various parts of the goggles, such as the frame, bridge, and temples. Additional video monitoring and recording devices, e.g, a mini Bluetooth/Wi-Fi camera, can be installed in the belt or collar (**Fig. 9**). Various local mirco-USB, secure digital (SD), or cloud-based [3], [10] storage devices and wireless access points (WAPs) can be installed in the earmuffs (**Fig. 10**). **Fig. 11** shows a hard or soft hat-mountable device, such as the WLAN, GPS, and LTE-capable Verizon JetPack[R] by ZTE [12], as a WAP for direct communication with an Intranet, e.g., body-area small cell or BASC, and as a gateway to the Internet for additional value-added BSNO services.

Devices that contain ECG sensors (Fig. 5) and wearable pulse meters (Fig. 6) are useful for lifestyle monitoring and vital-organ monitoring. Smart goggles and camera-embedded belt or collar are integral components of both types of black box services discussed in section 2. Bluetooth and Wi-Fi-based peer-to-peer (session and transaction) services can be offered via the earmuffs (Fig. 10) and/or hat-mounted virtual SSID capable WAP devices (Fig. 11). Device ID login, service ID log in, and lightweight key-based encryption are required for session and transaction. Message Queuing Telemetry Transport (MQTT) [13] is gaining popularity as a lightweight publish/subscribe model for message transport between machines and smart objects. Some simple extensions to MQTT for smart BSNO services may be very useful.

5 Architecture of Network Containing Smart Body Sensor Network Objects

Our network contains lightweight User Datagram Protocol (UDP) over IP and Transmission Control Protocol (TCP) over IP [1], [10] pipes, which lie over the underlying LTE, Wi-Fi, Bluetooth, and IEEE 802.15.4 [1]–[10], [17]–[21] network.

Fig. 12 shows a high-level architecture of such a network. Open server-side and open client-side APIs are used, and no specialized APIs are needed. Embedded web services using lightweight versions of protocols such as HTTP, XML, JSON, and Constrained Application Protocol (CoAP) [19] are used, depending on the footprint, power budget, and capability requirements. Applications and services based on Vital-Organ Moni-

▲**Figure 5. Wearable ECG sensor with RFID tag (indicated by "R") [14].**

▲Figure 6. Wearable pulse meter with RFID tag [15].

▲Figure 7. Front-and-backside wearable smart goggles (e.g., Google™ glass) with RFID tag.

▲Figure 8. Augmented realty goggles with RFID tag [16].

▲Figure 9. Miniature Bluetooth and Wi-Fi camera embedded into a belt or collar with RFID tag.

toring Cluster (VMC) run seamlessly and have low memory and processing overhead. These applications and services are used

▲Figure 10. Storage (mUSB and SD for PBB) and WAP mountable earmuffs with RFID tags.

▲Figure 11. Hat-mountable Verizon JetPack^{R} (by ZTE) with WLAN, GPS, and LTE.

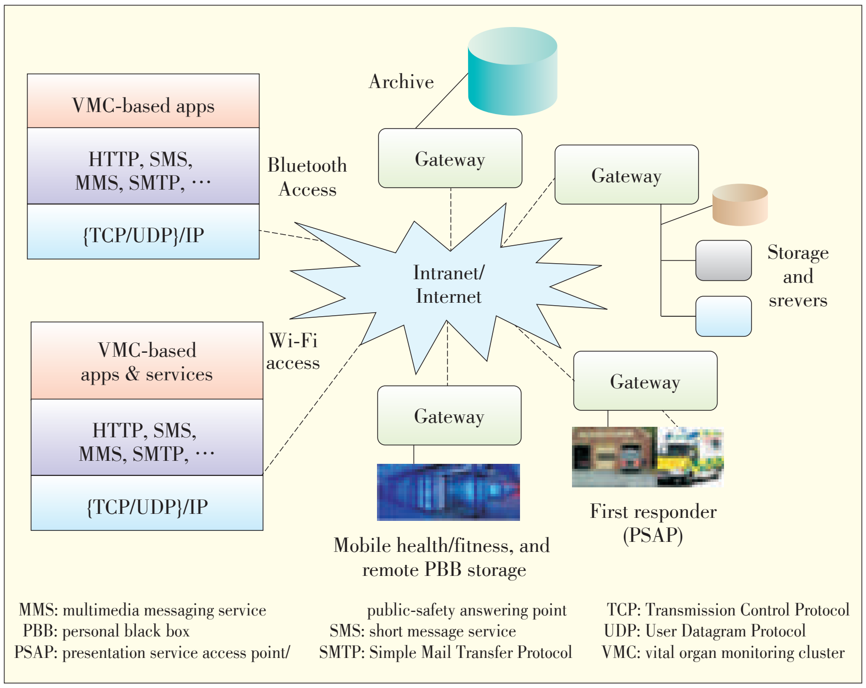
for smart body sensor object networking.

Although there is a universal definition of vital organs, in some special cases, it may be necessary to identify the vital organ in context. For example, the lungs can be considered more vital to patients with breathing difficulties.

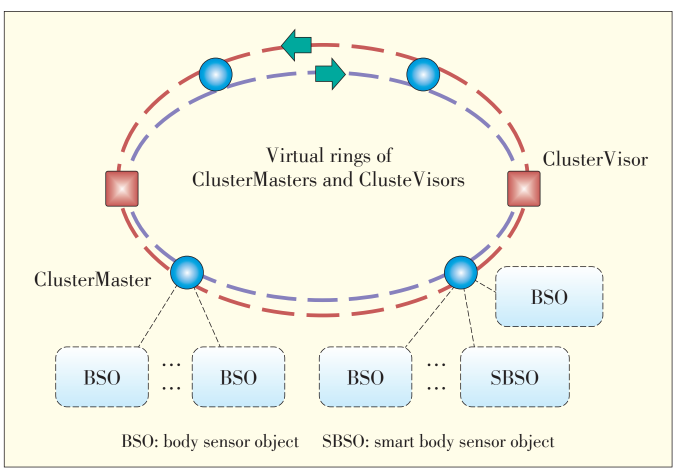
In addition to using an active RFID tag [1], [10], each BSO has another identifier for privacy and security reasons. Based on the pre-specified, pre-programmed interface, each BSO continuously or periodically logs sensed data in a certain format, e.g., comma-separated value (CSV) format. A BSO may also receive input data from secondary and tertiary BSOs that may be members of the same BSO cluster group (via ClusterMaster or ClusterVisor (**Fig. 13**). The stored log data is processed in real-time to locate anomalies, i.e., threshold crossing and correlated events, and then uploaded to be archived or to replenish stored information. For example, a refined version of MQTT [13] can be used for automated local and remote status updating and generation of triggers (e.g., alarm, call to ambulance).

Further analysis is done using clusters to discover abnormalities in the monitored information streams (**Fig. 14**).

Using the LTE uplink, information relating to anomaly and abnormality can be sent securely as top-secret information to



▲**Figure 12. High-level architecture of a network that uses body sensor network objects.**



▲**Figure 13. Architecture for clustering and virtual-ring-based communication between smart body sensor objects.**

remote databases. This top-secret information can be used by ambulance officers, specialists, and service providers for remote storage and appropriate preventative or corrective action.

6 Monitoring Models based on Body Sensor Network Objects

Depending on the state of the body organ being monitored, the state of a sensor may be idle (I) or active (A). In general, a simple two-state Markov chain model (**Fig. 15**) can be used to model the monitoring state of a body sensor device. By solving the simple two-state model with appropriate simplifying assumptions, the probabilities of idle state P_I and active state P_A is given as

$$P_I + P_A = 1 \quad (1)$$

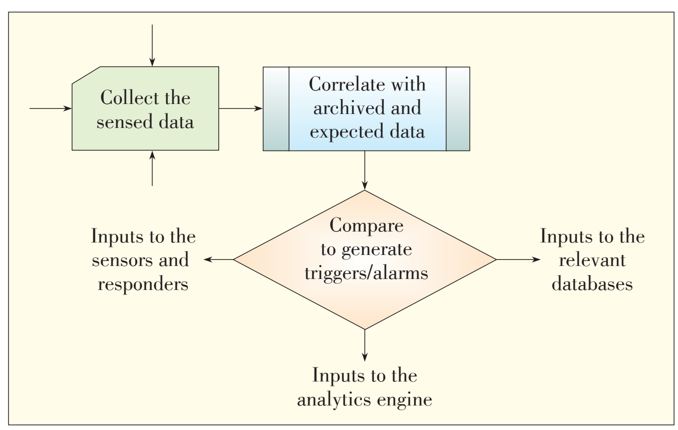
where $P_I = \frac{1}{1+(1/p_{AI})}$ and $P_A = \frac{1}{1+(1/p_{IA})}$.

Note that p_{AI} and p_{IA} are the transition probabilities, from active to idle and from idle to active. Depending on the use case, the probability of a sensor being active may be higher than the expected average of 50%. For example, during physical exercise, certain organs may need to be monitored more frequently in order to initiate preventative actions. Similarly, a sensor may remain idle for a long time depending on the context and use case.

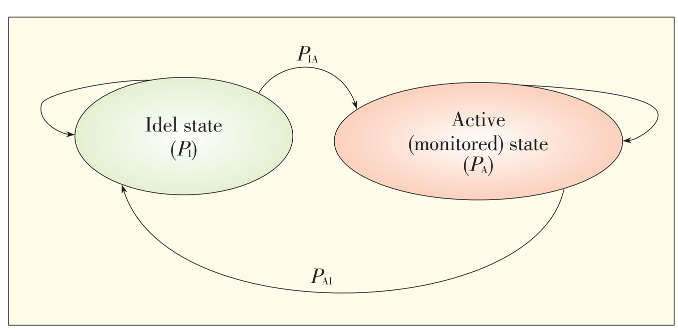
Fig. 16 shows a more realistic three-state monitoring model. The additional state here is the monitor-worthy (MW) state, and the probability of being in that state is given by P_{MW}. The MW state can be considered a transitional state between idle and active.

By solving a three-state Markov chain model, we find that

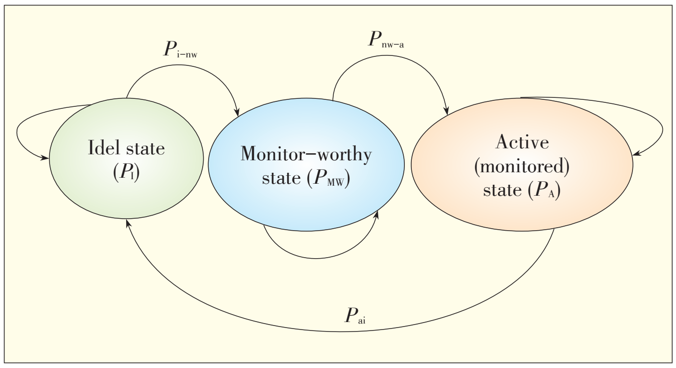
$$P_I + P_{MW} + P_A = 1 \quad (2)$$



▲**Figure 14. Steps for collecting and processing data collected by body sensors.**



▲**Figure 15. A simple two-state monitoring model based on body sensors.**



▲**Figure 16. A more realistic three-state monitoring model based on body sensors.**

where $P_I^{-1} = \left[1 + p_{IM}\left(\frac{1}{p_{MA}} + \frac{1}{p_{AI}}\right)\right]$,

$P_{MW}^{-1} = \left[1 + p_{MA}\left(\frac{1}{p_{IM}} + \frac{1}{p_{AI}}\right)\right]$, and $P_A^{-1} = \left[1 + p_{AI}\left(\frac{1}{p_{IM}} + \frac{1}{p_{MA}}\right)\right]$.

The probability of transition from idle to MW is denoted p_{IM}; the probability of transition from MW to active is denoted p_{MA}; and the probability of transition from active to idle is denoted p_{AI}. The probability that a sensor will be in any of three states is approximately 33% when the organ is functioning normally (hence of the sensor). However, depending on the context, situation, and use case, a sensor may remain in MW state more often than in the idle state.

7 Results of Experiments on Body Sensor Network

In this section, we present the preliminary results from a simple experimental setup where only a limited number of sensors are utilized. The basic monitoring, computing, and communications capability related features of the sensors along with some introductory results are presented in **Tables 1-4** and **Fig. 17**.

Data traffic appears in the cluster-master or cluster-visor and virtual ring (Fig. 12) in the pattern of micro-burst, peer-to-peer, or streams, depending on the type of monitoring described in Tables 1-4 and Fig. 17. For rapid prototyping and

▼**Table 1. ECG sensor's features as SBSO**

SBSO (device)	Desirable features for SBSON support					
	Virtualization	Message size	Routing	Encryption	Transmission	Host(s)
ECG sensor	Yes	~1024 bits	One-hop	Yes	~30 s; two-path	Dual

ECG: electrocardiogram SBSO: smart body sensor object SBSON: smart body sensor object network

▼**Table 2. Pulse meter's features as SBSO**

SBSO (device)	Desirable features for SBSON support					
	Virtualization	Message size	Routing	Encryption	Transmission	Hosts
Pulse meter	Yes	~512 bits	One-hop	Yes	~60 s two-path	Dual

SBSO: smart body sensor object SBSON: smart body sensor object network

▼**Table 3. Collar/belt-embedded camera features as SBSO**

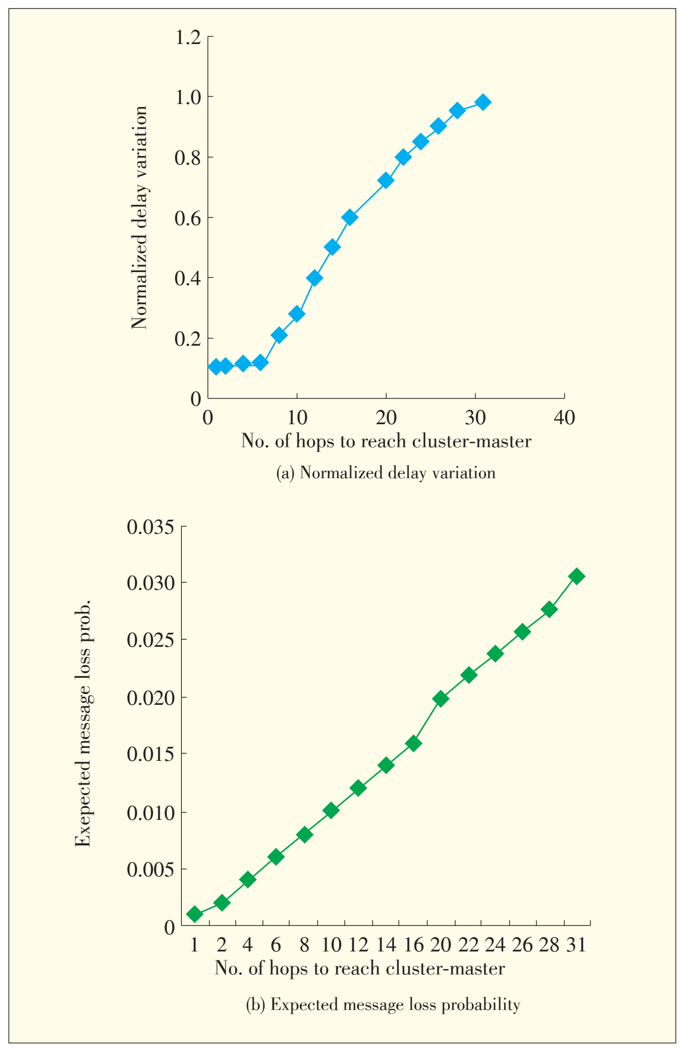
SBSO (device)	Desirable features for SBSON support					
	Virtualization	Message size	Routing	Encryption	Transmission	Hosts
Camera	Yes	Peer /streams (2-20 kilobits)	One-hop	May be required	Soft real-time (audio optional)	May need dual

SBSO: smart body sensor object SBSON: smart body sensor object network

▼**Table 4. Smart Goggle features as SBSO**

SBSO (device)	Desirable features for SBSON support					
	Virtualization	Message size	Routing	Encryption	Transmission	Host
Goggles	Yes	Peer /streams (5 to 15 frames per second)	One-hop	May be required	Hard real-time (with audio)	May need dual

SBSO: smart body sensor object SBSON: smart body sensor object network



▲**Figure 17. Preliminary dealy variation and loss probability.**

developing adaptive service-focused sensor apps, embedded LUA [22] can be used. The currently available eLUA modules [23] can be quickly customized and deployed in order to dynamically satisfy the analytical, operational, and environmental requirements.

Although no significant loss of monitoring data was observed using this simple setup, there was occasional variation in delay when the cluster-master or cluster-visor received monitored data. This can be attributed to the switching of the mode of transmission from Bluetooth or PAN to VLAN or Wi-Fi, and vice-versa. It may be possible to overcome this by segmenting the network, assigning one mode to each segment for communication to the cluster-master, or adding more cluster-masters to the ring.

The results in Fig. 17a and 17b are based on 32 SBSOs in a 32 virtual-hop ring with two cluster-masters and two cluster-visors.

We assume that:

- One in one thousand transmitted messages over a virtual hop (of the ring) can be corrupted or lost for various reasons (including mobility and interference).
- Cluster-master and cluster-visor has 32 MB or more of storage for each of the SBSOs in the ring for storing compressed and encrypted information.
- A message is considered obsolete (hence lost) once it has traversed 32 virtual hops (i.e., traversed all of the nodes of the virtual ring but failed to reach the cluster-master/-visor.

The channels with different transmisison technologies transmit information in different directions over the virtual channels, e.g., PAN channels are configured to transmit information in clockwise direction over the ring, and Bluetooth channels transmit information in counetr-clock-wise direction).

8 Conclusion and Future Work

We have discussed SBSOs, including their networking and internetworking in the context of a set of use cases. These use cases are lifestyle monitoring, vital-organ monitoring, and collecting data and information for black box services.

To make sensors smart, we imbued them with virtualization, simple predictive analytics, and proactive computing and communication. We pointed out general usage and deployment etiquette in Fig. 4b and regulatory implications in the context of the use cases mentioned. [26]-[28]

We presented simple two-state and three-state models of monitoring by sensors as well as preliminary results from a simple experimental setup where only a limited number of sensors were used.

We also discussed the use of virtualized ad-hoc networking using cluster-master or cluster-visor.

The following are future research directions that build on what we have presented here:

- Personalization of sensors. Probes and capabilities of sensors can be personalized according to a person's needs.
- Session- and transaction-specific authorization and authentication of the sensors. This refers to the prevention of unauthorized local or remote uploading of the monitored information for any session or transaction to or from the sensors.
- Self-destruct or beacon emitting sensors. This protects sensors against hacking or theft. A kill-switch can be embedded in the sensors in order to prevent any unauthorized use of lost or stolen sensors.
- Mechanisms to improve reliability, survivability, and sustainability of the sensor networks. This refers to the development and invocation of reliability-improving mechanisms in a virtualized environment for both body sensors and sensor network objects.
- Self-healing and self-monitoring smart sensors. This refers to developing implementable body sensors that have self-monitoring, self-calibrating, and self-healing components.
- Sensors powered by solar, body heator friction. This refers to developing low-power body sensors that can operate using a remote charger, solar power, body heat, or friction created by limbs or organs. A recent invention is a remote charger that powers wireless medical implants deep inside the body [24].
- Miniaturization of the SBSOs in order to embed them seamlessly in body and everyday wearable garments. This refers to miniaturization of the body sensors so that these can be embedded in everyday wearable items like garments, shoes, ornaments, etc. A recent experiment utilized a miniature implantable device to treat heart failure patients with central sleep apnea [25].
- Mobile cloud [3] based operations and management of BSOs: This refers to seamless use of virtualization and automation for management, operation and orchestration of body sensor objects across multiple domains for mobile users.

References

[1] A. S. Esguevillas, B. Carro, B. Khasnabish, and A. Gupta, "Guest editorial: Applications and support technologies for mobility and enterprise services," *IEEE Wireless Communications Magazine*, vol. 16, no. 3, pp. 6-7, Jun. 2009.

[2] H. Tschofenig and J. Arkko. (2012, Apr.). Report from the smart object workshop. [Online]. Available: http://tools.ietf.org/rfc/rfc6574.txt

[3] B. Khasnabish. (2012 Sept.). Mobile cloud for personalized any-media services. *ZTE Communications* [Online]. *10(3)*, pp. 47-54. Available: http://tinyurl.com/bhumip-mobile-cloud

[4] J. Gilger and H. Tschofenig. (2012 Oct.). Smart object security workshop report. [Online]. Available: http://tools.ietf.org/id/draft-gilger-smart-object-security-workshop-00.txt

[5] IETF. (2014). Light-Weight Implementation Guidance (LWIG) activities. [Online]. Available: http://datatracker.ietf.org/wg/lwig/charter

[6] IETF. (2014). Network Virtualization Overlays (NVO3) activities. [Online]. Available: http://datatracker.ietf.org/wg/nvo3/charter

[7] IETF. (2014). System for Cross-Domain Identity Management (SCIM) activities. [Online]. Available: http://datatracker.ietf.org/wg/scim/charter

[8] IETF. (2014). Mobile Ad-hoc Networks (MANET) activities. [Online]. Available: http://datatracker.ietf.org/wg/manet/charter

[9] IRTF. (2014). Software Defined Networking Research Group (SDN-RG) activities. [Online]. Available: https://irtf.org/sdnrg

[10] T. Plevyak and V. Sahin, *Next Generation Telecommunications Networks, Services, and Management*. New Jersey: Wiley-IEEE Press Book, Apr. 2010.

[11] C114. (2014 Apr.). Worldwide wearable computing device shipment prediction.

[Online]. Available: http://www.cn-c114.net/583/a830917.html
[12] ZTE Corporation. (2014). Verizon JetPack® 890L. [Online]. Available: http://www.zteusa.com/verizon-jetpacktm-890l.html
[13] MQTT. (2014). Message Queuing Telemetry Transport (MQTT) for lightweight publish/subscribe messaging transport. [Online]. Available: http://mqtt.org/
[14] NICT. (2014). Wearable ECG sensor image. [Online]. Available: http://www.nict.go.jp
[15] Clinical Guard. (2014). Wearable pulse meter image. [Online]. Available: http://www.clinicalguard.com
[16] Recon Instruments. (2013). Augmented reality Goggle image. [Online]. Available: http://www.reconinstruments.com
[17] IETF. (2014). IPv6 over Low Power WPAN (6LOWPAN) activities. [Online]. Available: http://datatracker.ietf.org/wg/6lowpan/charter
[18] IETF. (2014). Routing Over Low Power and Lossy Networks (ROLL) activities. [Online]. Available: http://datatracker.ietf.org/wg/roll/charter
[19] IETF. (2014). Constrained RESTful Environments (CORE) activities. [Online]. Available: http://datatracker.ietf.org/wg/core/charter
[20] IETF. (2014). Transport Layer Security (TLS) activities. [Online]. Available: http://datatracker.ietf.org/wg/tls/charter
[21] IEEE. (2014). 802.15 WPAN™ Task Group 4 (TG4). [Online]. Available: http://www.ieee802.org/15/pub/TG4.html
[22] eLua. (2014). Service-focused sensor apps using Embedded LUA (eLUA). [Online]. Available: http://www.eluaproject.net
[23] eLua. (2014). Service-focused Embedded LUA (eLUA) modules. [Online]. Available: http://www.eluaproject.net/overview/status
[24] MNT. (2014, May 21). Invention powers wireless medical implants deep inside the body. [Online]. Available: http://www.medicalnewstoday.com/articles/277110.php
[25] MNT. (2014, May 20). Implantable device treats heart failure patients with central sleep apnea. [Online]. Available: http://www.medicalnewstoday.com/articles/277032.php
[26] IEEE. (2014 May). Body Sensor Networks (BSN) Conf. [Online]. Available: http://www.bsn2012.org
[27] IEEE. (2013 May). Body Sensor Networks (BSN) Conf. [Online]. Available: http://toc.proceedings.com/19075webtoc.pdf
[28] IEEE. (2014 Jun.). Body Sensor Networks (BSN) Conf. [Online]. Available: http://www.bsn2014.org

Manuscript received: 2014-03-24

Biography

Bhumip Khasnabish (bhumip.khasnabish@ztetx.com), PhD, is a senior member of the IEEE and an emeritus distinguished lecturer of the IEEE Communications Society. He initiated cloud and data center activities in IETF and is vice-chair of DMTF NSM WG. He previously co-chaired the ATIS IPTV Interoperability Forum (IIF) and founded and chaired ATIS NG-CI TF and MSF Services WG. He is currently senior director in the Strategy Planning and Standards Development Department, ZTE USA. His research interests include next-generation networking; platform and services that use virtualized computing, communication, and service entities; tighter cross-layer communications; software-defined networking; and system configuration/service automation.

Dr. Khasnabish has authored numerous articles, book chapters, books, and patents in areas related to converged services and next-generation networking. His latest book, *Implementing Voice over IP* (Wiley-IEEE Press, 2003) is currently in its second printing. Dr. Khasnabish has co-edited *Multimedia Communications Networks: Technologies and Services* (Artech House, 1998) and many special issues of *IEEE Network*, *IEEE Wireless Communications*, *IEEE Communications*, and the *Journal of Network and Systems Management*.

Smartphones to Account for 2/3 of World's Mobile Market by 2020, Says New GSMA Intelligence Study

11 September 2014, Hong Kong: Smartphones will account for two out of every three mobile connections globally by 2020, according to a major new report by GSMA Intelligence, the research arm of the GSMA. The new study, "Smartphone forecasts and assumptions, 2007 – 2020", finds that smartphones account for one in three mobile connections today, representing more than two billion mobile connections. It forecasts that the number of smartphone connections will grow three-fold over the next six years, reaching six billion by 2020, accounting for two-thirds of the nine billion mobile connections by that time. Basic phones, feature phones and data terminals will account for the remaining connections. The study excludes M2M from the connections totals.

"The smartphone has sparked a wave of global innovation that has brought new services to millions and efficiencies to businesses of every type," said Hyunmi Yang, Chief Strategy Officer at the GSMA. "As the study released today shows, smartphones will be the driving force of mobile industry growth over the next six years, with one billion new smartphone connections expected over the next 18 months alone.

The developing world overtook the developed world in terms of smartphone connections in 2011 and today accounts for two in every three smartphones on the planet, according to the new study. It is predicted that by 2020, four out of every five smartphone connections worldwide will come from the developing world.

Asia Pacific today accounts for about half of global smartphone connections, even though smartphone penetration in the region is currently calculated at below 40 per cent. The Asia Pacific total is boosted by the inclusion of China, the world's largest smartphone market, with more than 629 million smartphone connections.

In many developed markets, smartphone penetration is approaching the 70 to 80 per cent "ceiling" at which growth tends to slow. According to the report, smartphone adoption is forecast to reach 75 per cent in Europe and North America by 2020. Smartphone growth in these two regions has slowed in recent years; smartphone connections grew by 35 per cent in North America and by 39 per cent in Europe between 2010 and 2013, compared to growth rates of over 80 per cent during the period in Asia Pacific and Latin America. (Source: c114)

DOI: 10.3939/j.issn.1673-5188.2014.03.006
http://www.cnki.net/kcms/detail/34.1294.TN.20140818.1654.001.html, published online 18 August, 2014

E-Healthcare Supported by Big Data

Jianqi Liu[1], Jiafu Wan[2], Shenghua He[1], and Yanlin Zhang[1]
(1. School of Information Engineering, Guangdong Mechanical and Electrical College, Guangdong 510515, China;
2. School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou 341000, China)

Abstract
The era of open information in healthcare has arrived. E-healthcare supported by big data supports the move toward greater transparency in healthcare by making decades of stored health data searchable and usable. This paper gives an overview the e-healthcare architecture. We discuss the four layers of the architecture—data collection, data transport, data storage, and data analysis—as well as the challenges of data security, data privacy, real-time delivery, and open standard interface. We discuss the necessity of establishing an impeccably secure access mechanism and of enacting strong laws to protect patient privacy.

Keywords
healthcare; wireless body network; big data; disease prediction; remote monitoring; medical data

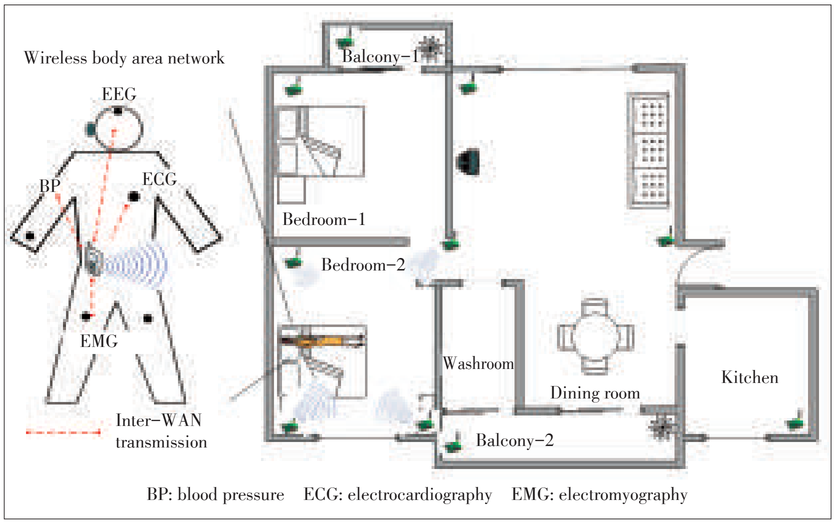
1 Background

In China, more than 10% of the population is over sixty, and by 2015, the number of elderly in China will be 200 million. A rapidly aging population is one of the long-term effects of the one-child policy introduced in the 1970s. This trend could lead to a huge labor shortage by 2050 as well as increased demand for more paramedics and medical services. Many experts warn that healthcare for the elderly is an urgent issue. In America, healthcare expenses have been rapidly increasing over the past two decades; high medical costs is not a new phenomenon. In 2009, medical costs in the United States were 17.6% of GDP, nearly $600 billion more than expected. Escalating costs are starting to alter provider reimbursement trends. Risk-sharing models have started to replace many fee-for-service models in order to curb expenses and encourage judicious use of resources. The patient is shouldering more of the cost of medical services than ever before. However, in the United States, the elderly are less likely to be able to pay for these services. This problem is even more serious in China, where many older people so not have health insurance. If they fall chronically ill, their families will not be able to afford the bill. Also, lack of real-time monitoring is an issue. Real-time physiological data is very important for treating heart disease in particular, but collecting such data is not easy because heart disease is usually characterized by fleeting symptoms. Similar other diseases require real-time monitoring. There are some issues in the Chinese healthcare system, such as slow response and isolated clinical data; however, advances in wireless body-area networking (WBAN) [1], [2] and big data [3] introduce new ways to address these issues. A modern e-healthcare system takes full advantage of big data [4] and heterogeneous networks, which include WBAN, 3G, and 4G technologies. An e-healthcare system collects data on the physical and mental health of a patient and also collects data from clinical records. It then uses data-mining algorithms to provide well-reasoned, useful information for patient care. E-healthcare has three main benefits:

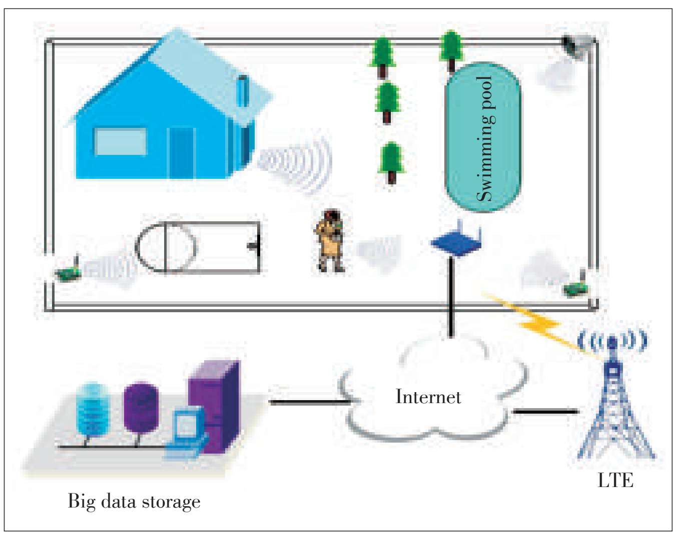
- It eliminates redundant treatment by providing the doctor with a more comprehensive medical profile of the patient. This helps the doctor precisely target their treatment.
- It provides continuous, timely monitoring and diagnosis. A WBAN can monitor the vital signs of patient and transmit this information in real time to a data center via cellular network or the Internet. This enables timely diagnosis.
- It reduces costs. The patient may be eligible for discount medicines by authorizing the medical data to be used by a pharmaceutical company.

Here we consider two e-healthcare scenarios. In scenario one, (**Fig. 1**), an elderly person with health problems is sleeping at home. Sensors on the body continuously monitor the person's vital signs, motor activity, social interaction, sleep patterns, and other health indicators. The data is sent to a big-data storage server through the hybrid network so that a doctor can assess the data and make an accurate diagnosis [5], [6].

In scenario two (**Fig. 2**), an elderly person takes a walk in the garden. Sensors are deployed on the balcony and in the yard to detect sunlight and weather conditions outdoors. The collected data is so large and complex that it cannot be processed using conventional database management and data-pro-



▲Figure 1. Body sensors collect vital signs in an indoor home environment.



▲Figure 2. An elderly person takes a walk in the garden. Sensors collect vital signs and environmental information and transmit this to a big-data storage server via a hybrid network.

cessing tools. Big-data applications are therefore introduced.

The e-healthcare scenarios described here benefit the patient as well as the provider, medical institution, pharmaceutical company, and government. E-healthcare has four main requirements: usability and comfort, data acquisition, real-time communication, and big-data processing. The key design parameters for an e-healthcare system are:

- small, light, wearable sensor modules
- perception of both indoor and outdoor environments by the WSN
- low power consumption (to avoid frequent changing of batteries)
- real-time transmission
- data-analysis capabilities
- disease prediction model.

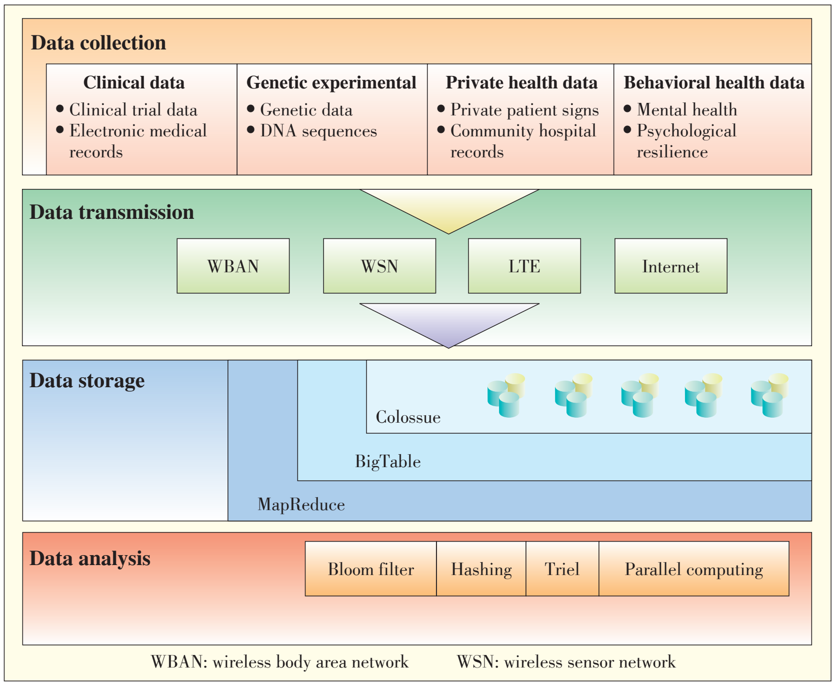
E-healthcare is designed to be discrete, usable, comfortable, and intelligent.

2 Architecture

In the health sciences, scientific method is based on experimentation or clinical data, but the limiting factor is lack of relevant data that either supports or refutes the initial hypothesis. Patients cannot benefit from their historical medical and environmental data. With the development big-data techniques, huge amounts of data are derived from patients, doctors, research institutions, and pharmaceutical companies. The e-healthcare system needs to take advantage of these massive amounts of data and provide the right intervention to the patient at the right time. It should offer personalized care to the patient. According to Peter Groves et al., an e-healthcare system should achieve the following goals [7]:

- Right lifestyle. Patients can maintain a positive attitude in their treatment, including complying with the doctor´s treatment guidelines and taking active measure to prevent disease. With advice given by e-healthcare, people can make proper lifestyle choices that help them remain healthy.
- Right medical care. Patients should be offered the most reasonable treatments when sick. In China, some doctors tend to over-treat or use expensive drugs in order to obtain a higher income. E-healthcare can provide treatment examples for the same disease. Remote monitoring and diagnosis provided by e-healthcare is a particularly good choice for outpatients. If a patient is in crisis, e-healthcare sends a warning to the doctor for quick intervention.
- Right medical workers. Patients should always be treated by high-performing professionals that are best matched to treat the disease. However, patients often cannot get detailed information about a doctor and are unable to make appropriate choices. The e-healthcare system should provide comprehensive information about medical staff so that patients can make informed decisions.
- Right value. Medical workers, payers and other institutions will continuously enhance e-healthcare value while preserving or improving on quality. E-healthcare can offer advice about the best treatments; it can avoid overtreatment; it offers real-time monitoring; and it involves multiple measures for ensuring cost-effectiveness. It can eliminate fraud, waste, and abuse in the system.

An e-healthcare architecture supported by big data has four layers: data collection, data transport, data storage, and data analysis (**Fig. 3**). The data-collection layer includes wireless sensors, for collecting environmental data such as temperature and humidity, and a GPS receiver, for determining the location of the patient. The data-transport layer converges raw data



▲Figure 3. E-healthcare architecture.

from the wireless sensor network (WSN), WBAN to router, and sends this data to a data center via traditional Internet or LTE cellular network. In the data storage and analysis layer, the patient's health information is collected and a prediction is made on possible signs of disease.

3 Data Collection

3.1 Data Classification

E-healthcare's innovation mostly depends on the patients' medical data. Some research institutions use private data derives from medical experiments. But many hospitals still capture patients' medical data to make further research. According to the data sources, the data can be categorized as clinical trial, genetic experimental, private health, mental health, and other according to the source of the data.

Clinical trial data is derived from public papers or electronic medical records [8]. In clinical trials, the quality of the data collected depends first and foremost on the quality of instruments used. No matter how much time and effort go into conducting a trial, if the correct data points are not collected, it may not be possible for an e-healthcare system to do a meaningful analysis. Therefore, the design, development, and quality assurance of medical instrumentation must be given the utmost attention. Clinical trial data is structured and very easy to put into big-data storage. As well as structured electronic medical records, there are also some unstructured clinical notes and medical images. These are also important data sources for an e-healthcare system.

Experimental data described here is private or proprietary data, usually derived from R&D conducted by pharmaceutical companies or government institutions [9]. This data is usually confidential and not publically available. It includes genetic data (DNA sequences). If the system is secure, limited sharing of data will improve the universality and veracity of the data system.

Private health data includes data on vital signs and environment collected by sensors on the patient's body and the personal data of a patient collected from a hospital.

Mental health may include an individual's ability to enjoy life, balance life activities, and achieve psychological resilience. The state of a person's mental health can be collected by a trained psychologist. Statistics from the World Health Organization show that nearly half the world´s population is affected by mental illness, which affects people's self-esteem, relationships and ability to function in everyday life. Long-term psychological illness impacts physical health. Therefore, mental health data is important in an e-healthcare system [10].

There is also other data, such as epidemiology data, environmental data and location data, which affects the health of a patient [11].

3.2 Sensors

A sensor node mainly comprises the physiological and environmental signal sensor and the radio platform that sensors are connected to. The general function of a sensor is to collect analog signals that correspond to the human physiological condition. The analog signals are received by a corresponding radio-equipped board, where they are then digitized. The digital signals are then forwarded by the router to the data center, where they are stored.

A blood pressure sensor is a non-invasive sensor for measuring systolic and diastolic blood pressures. The sensor automatically inflates an upper arm cuff at customizable time intervals to acquire readings and transmit the data to the storage server. The sensor was developed by Johns Hopkins University Applied Physics Laboratory and is based on the NIBP module from SunTech Medical. It takes a reading every five minutes [12]. It has a battery pack comprising four 9 V lithium batteries and can operate for ten hours. It shows a pulse rate accurate to ±3 beats per minute and blood pressure accurate to ±3 mmHg [13].

A CO_2 sensor measures gaseous carbon dioxide level and oxygen concentration in human respiration. In [14], a wireless, passive carbon-nanotube gas sensor is introduced. Multiwall carbon nanotubes (MWNTs) remotely receive data on carbon dioxide, oxygen, and ammonia based on the measured changes in MWNT permittivity and conductivity with gas exposure.

An electrocardiographic (ECG) sensor records the heart's electrical activity. Healthcare providers use it to help diagnose heart disease. In order to obtain an ECG signal, several electrodes are attached at specific sites on the skin, and the difference in electrical potential between these electrodes is measured [15].

A humidity and temperature sensor measures the temperature of the human body as well as the humidity and temperature of the surrounding environment [16].

A pulse oximetry (SpO2) sensor measures the oxygen saturation and heart rate through a non-invasive probe. The SpO2 sensor board developed by Smiths Medical is used in ambulances and is very accurate. According to the manufacturer's specifications, it has an SpO2 accuracy of ±2% and heart rate accuracy of ±2 bpm. The sensor board was chosen for the AID-N project.

A camera sensor may be a complementary metal-oxide-semiconductor (CMOS) active-pixel sensor, such as OV9650, that is embedded in a PDA, cellphone, or other device. Video recordings have been used to treat Parkinson's disease. An expert examines the video recordings and provides clinical cores representing the severity of the tremor, dyskinesia, and bradykinesia.

A vital signs sensor benefits from the development of wireless communications; in particular, miniaturization and energy efficiency in embedded computing has improved significantly. The sensor node is therefore becoming much smaller.

4 Data Transport

In e-healthcare, data is transmitted via heterogeneous networks. Physiological and mental health data is relayed via the WBAN; environmental data is relayed via wireless sensor networks; and experimental or clinical data is relayed via traditional Internet. In situations where long-distance transmission is required and there is no wired network, LTE should be deployed.

WBAN is one of the best technologies for building unobtrusive, scalable, robust wearable health monitoring systems. A WBAN for health monitoring comprises multiple sensor nodes. Each node is typically capable of sensing and processing one or more physiological signals, caching them, and transmitting the data to a big-data storage server.

A WSN comprising spatially distributed autonomous sensors monitors environmental conditions and passes the data via a wireless network to the storage server. Modern networks are bi-directional, meaning that sensor activity can be controlled. Today such networks are mainly used in industrial and consumer applications, such as monitoring and control of industrial processes and machine health.

LTE is 4G communication technology that enables peak download rates of up to 299.6 Mbit/s and peak upload rates of up to 75.4 Mbit/s depending on the user equipment category (with 4 × 4 antennas and 20 MHz of spectrum). In the case of a moving patient, LTE can easily transmit the vital signs to the server and overcome the defects of a traditional network [17].

5 Data Storage

Mass medical data requires an appropriate storage mechanism. Google products are stable solutions and can be classified into three levels. From the bottom up these are: basic file system, such as Google file system (GFS) or Colossus; management database, such as BigTable; and programming model, such as MapReduce [18].

Google's Colossus is second-generation GFS and an expandable distributed file system that supports large-scale, distributed, data-intensive applications. It overcomes problems such as a single point failure and poor performance with small files. Colossus is the foundation of upper-level applications. In addition, open-source file systems, such as HDFS and Kosmosfs can be acquired [19].

BigTable is a column-oriented, distributed, structured data storage system designed to process large-scale (petabyte) data of thousands of commercial servers. The basic data structure of BigTable is multidimensional sequenced mapping with sparse, distributed, persistent storage. BigTable is based on many fundamental components of Google, including Colossus, cluster management system, SSTable file format, and Chubby.

MapReduce [20] is a simple but powerful programming model for large-scale computing. It uses a large number of commercial PC clusters for automatic parallel processing and distribution. The MapReduce computing model only has two functions—Map and Reduce—both of which are programmed by the user. The Map function processes input key-value pairs and generates intermediate key-value pairs. MapReduce then combines all the intermediate values related to the same key and transmits them to the Reduce function, which further compresses the value set into a smaller value set.

6 Data Analysis

In the healthcare industry, medical data is important role predicting diseases, and some researchers use the genome data (DNA) for this purpose. When medical experts identify genes in DNA that are markers for disease, a person can make appropriate lifestyle or other changes to lower the risk of disease. Genome data has used to predict heart disease and brain disorders. For inherited diseases, identifying a parent who is a carrier but does not express the disease can help parents make informed choices regarding pregnancy. Some researchers use data-mining to predict diabetes and breast cancer according to the health profile of the individual. Some researchers use environmental data or information on epidemics to predict diseases such as the flu. In the big-data era, people are concerned about how to rapidly extract important information from mass data in order to benefit both enterprises and patients. At present, the main methods for processing big data are [18]: bloom filter, hashing, indexing, triel, and parallel computing.

A bloom filter comprises a series of hash functions. The principle of a bloom filter is to store the hash values of data, rather than the data itself, by using a bit array. This is, in essence, a bitmap index that uses hash functions for lossy data compression and storage. It is highly efficient in terms of query speed and space-saving, but it also has some disadvantages, such as misrecognition and deletion.

Hashing is a method is a method for transforming data into shorter fixed-length numerical values (index values). It has the advantages of rapid reading, writing, and query speed, but it is difficult to find a sound hash function.

Indexing is an effective way of reducing disk reading and writing and improving insertion, deletion, modification, and query speeds in both traditional relational databases, which manage structured data, and other technologies that manage semi-structured and unstructured data. However, with indexing there is additional cost for storing index files, which need to be maintained dynamically when data is updated.

Triel (also called trie tree) is a variant of hash tree. It is mainly used for rapid retrieval and word frequency statistics. The main principle of triel is to use common prefixes of character strings to reduce comparison of character strings to the greatest possible extent and increase query efficiency.

Parallel computing, unlike traditional serial computing, involves the simultaneous use of several computing resources to complete a computing task. The basic principle of parallel computing is to decompose a problem and assign the parts of the problem to several separate processes to be independently dealt with. This is also called co-processing.

We can envision a thorough healthcare system in which medical devices, wearables, diagnostic tools, and analytics empower patients and their families to better care for themselves. A disease-prediction model is essential to the e-healthcare system. Real-time data alerts enable the doctor or healthcare provider to intervene when necessary. A developed disease-prediction model first and foremost needs to provide accurate, internally and externally validated probabilities of specific health conditions or outcomes in a patient. Such models must guide the doctor's decision-making process and the patient's behavior. This will improve the outcome for the patient and reduce the cost of care.

6.1 Breast Cancer Prediction Model

Researchers, clinicians, and the public are increasingly interested in statistical models designed to predict the occurrence of cancer [21]. Susan M. Domchek et al. use a cancer risk prediction model to estimate the risk of breast cancer for women [21]. The model estimates the likelihood of breast cancer risk due to genetic susceptibility, such as BRCA1 or BRCA2 mutations. Recent developments have reinforced the clinical importance of breast cancer risk assessment. Tamoxifen chemoprevention and studies such as the Study of Tamoxifen and Raloxifene are available to women who are at increased risk of breast cancer [26]. In addition, specific management strategies are now defined for BRCA1 and BRCA2 mutation carriers. Risk-assessment can be used to determine the likelihood that a woman will develop breast cancer, or prior probability models can be used to determine the likelihood of BRCA1 or BRCA2 mutation. Several models may be needed to give the woman optimal counseling, provide the woman and her family with accurate, useful information, and make a sound clinical judgment.

6.2 Coronary Heart Disease Prediction Model

A computer simulation model was developed to project the future mortality, morbidity, and cost of coronary heart disease (CHD) in the United States [22]. The model contains a demographic-epidemiologic submodel that simulates the distribution of coronary risk factors and the conditional incidence of CHD in a demographically evolving population. It also contains a "bridge" submodel that determines the outcome of the initial CHD event as well as a disease history submodel that simulates subsequent events in a person who has previously had a CHD event. The effects of either preventive or therapeutic intervention on mortality, morbidity, and cost can be simulated for up of to a 30 years. The baseline projection is based on no changes in risk factors or efficacy of therapies after 1980. It shows how the aging of the population, especially the baby-boomer generation, naturally increases the annual incidence, mortality, and costs of CHD by about 40-50% by 2010. Unprecedented reduction of risk factors would be required to offset these demographic effects on the absolute incidence of CHD. However, forecasts could be inaccurate because of misplaced assumptions or poorly estimated baseline data, and the model awaits validation using actual future data.

7 Applications

Some healthcare leaders are already gaining value from healthcare aided by big data. The following two examples show how mass medical data can be used to reduce the medical costs.

7.1 HealthConnect

In 2002, Kaiser Permanente ceased construction of its own clinical information system and ultimately sought out another vendor, Epic Systems, to undertake this construction. The new system, called HealthConnect [23], [24], is one of the largest private electronic health systems in the world. The system integrates more than 611 medical offices and 37 hospitals, linking patients to their healthcare teams and private health data. This system has served 9.1 million members and helps patients refill 1.2 million prescriptions monthly. HealthConnect facilitates communication between members and doctors to help make getting well and staying healthy easy and convenient. HealthConnect share the latest findings and best practices that a comprehensive electronic health record can increase consumer convenience and satisfaction and provider efficiency, while maintaining clinical quality.

In addition, the mass data in HealthConnect helps clinicians with their research. Kaiser Permanente operates one of the largest non-university research programs in the United States. Kaiser Permanente clinicians and researchers have approximately 2000 studies in progress at any given time and publish 900 to 1000 articles annually. Accessing fully digitized health records is much cheaper and far less time consuming than accessing paper records when conducting research. With this mass-digitized data, Kaiser Permanente researchers have been able to explore a number of areas, including how vaccines, medications, and lifestyle affect the whole population. The research outcome is used to improve the healthcare.

7.2 IMS Health Disease Analyzer

The Disease Analyzer patient database contains data on diagnoses, prescriptions, risk factors, and laboratory results for approximately 5 million patients per year in Germany [25]. The database also contains data from various groups of specialist doctors and general practitioners. Data from approximately 3200 office-based doctors form the basis of this investigation. Data are delivered with OLAP based software application (Disease Analyzer Software) or data are used for ad-hoc studies which are sold with analytic services to customers.

8 Challenges

With the more widespread release of personal health information, the government, leading companies, and research institutions need to consider regulations about its use, as well as privacy protections and data security. To encourage data sharing and streamline the repetitive nature of granting waivers and data-rights administration, it may be better for data approvals to follow the patient, not the procedure. Further, data sharing could be made the default, rather than the exception. It is important to note, however, that as data liquidity increases, physicians and manufacturers will be subject to increased scrutiny, which could result in lawsuits or other adverse consequences.

8.1 Data Privacy

Personally identifiable information or other sensitive information including healthcare records, biological traits, such as genetic material is possible to be divulged in e-healthcare. The challenge in data privacy is to exchange data while protecting personally identifiable information. As heterogeneous information systems with differing privacy rules are interconnected and information is shared, policy appliances will be required to reconcile, enforce and monitor an increasing amount of privacy policy rules (and laws). There are two categories of technology to address privacy protection in commercial IT systems: communication and enforcement. As a user, you also need to protect yourself data privacy, on the internet you almost always give away a lot of information about yourself: Unencrypted e-mails can be read by the administrators of the e-mail server, if the connection is not encrypted (no https), and also the internet service provider and other parties sniffing the traffic of that connection are able to know the contents.

8.2 Data Security

The e-healthcare system can realize benefits from democratizing big data access. The researchers can more easily collaborate, engage in peer review and eliminate duplication of efforts. Researchers will also be able to more readily identify opportunities where they can contribute and collaborate. The system makes exposing and sharing data easy and relatively inexpensive. However, significant security concerns exist. A credentialing process could facilitate and automate this access, but there are complexities and challenges. Since providers, patients and other interested parties such as researchers need secure access; data access should be controlled by group, role and function. Finally, the security of the data once it leaves the cloud also needs to be assured. Big data can be used to identify patterns and irregularities indicating and preventing security threats, as well as other types of fraud.

8.3 Real-Time Communication

The healthcare is performance-critical applications, which require bounded delay latency. Whether the heterogeneous network has capability of providing bounded delay guarantees on packet delivery is very important to monitor. Since WSN and WBAN deal with real world, it is often necessary for communication to meet real-time constraints. In surveillance systems, for example, communication delays within sensing and actuating loops directly affect the quality of tracking. To date, few results exist for WSNs that adequately address real-time requirements. Tian He and John A Stankovic proposed a real time protocol SPEED to meet requirement. The redundancy of nodes is a solution for wireless network. The presence of multiple disjoint paths between nodes makes them robust to link and node failures.

8.4 Data Format and Standard Interface

The data of e-healthcare are derived from different institutions, the formats are different. These heterogeneous systems need to adapt to different formats and interfaces, affect the exchange of data. So we need an open standard for data formats along with open standardized interfaces, and the healthcare system can access each other more easily with various healthcare providers, hospital, and pharmaceutical company. This is a challenge for e-healthcare, the government or guild has a responsibility to build a specification for data exchange.

9 Conclusion

In order to achieve these goals, including right living, right care, right provider, right value, and right innovation, the big data technique have been introduced into e-healthcare. With the support of big data, the e-healthcare can resolve massive data storage, data management, and data analysis. If e-healthcare can protect the personal data security and privacy, the e-healthcare can reduce the healthcare cost by a large margin. The disease prediction can give patient appropriate strategies to lower the risk of disease.

Acknowledgement

The authors would like to thank the Natural Science Foundation of Guangdong Province, China (No. 9151009001000021), the Ministry of Education of Guangdong Province Special Fund Funded Projects through the Cooperative of China (No. 2009B090300341), the National Natural Science Foundation of China (No. 61262013), the Open Fund of Guangdong Province Key Laboratory of Precision Equipment and Manufacturing Technology (No. PEMT1303), and the Higher Vocational Education Teaching Reform Project of Guangdong Province (No.20130301011) for their support in this research.

References

[1] M. Chen, S. Gonzalez, V. Leung, and Q. Zhang *et al.*, "A 2G-RFID-based e-healthcare system," *IEEE Wireless Communications*, vol. 17, pp. 37–43, 2010. doi: 10.1109/MWC.2010.5416348.

[2] J. F. Wan, C. F. Zou, and S. Ulah *et al.*, "Cloud-enabled wireless body area networks for pervasive healthcare," *IEEE Network*, vol. 27, no. 5, pp. 56–61, 2013. doi: 10.1109/MNET.2013.6616116.

[3] S. LaValle, E. Lesser, and R. Shockley *et al.*, "Big data, analytics and the path from insights to value," *MIT Sloan Management Review*, vol. 21, 2014

[4] H. Cao, V. Leung, and C.Chow *et al.*, "Enabling technologies for wireless body area networks: A survey and outlook," *IEEE Communications Magazine*, vol. 47, pp. 84–93, 2009. doi: 10.1109/MCOM.2009.5350373.

[5] J. Liu, Q. R. Wang, and J. F. Wan *et al.*, "Towards key issues of disaster aid based on wireless body area networks," *KSII Transactions on Internet and Information Systems*, vol. 7, pp. 1014–1035, 2013.

[6] F. H. Messerli, B. D. Nunez, and H.O. Ventura *et al.*, "Overweight and sudden death: increased ventricular ectopy in cardiopathy of obesity," *Archives of Internal Medicine*, vol. 147, pp. 1725, 1987.

[7] P. Groves, B. Kayyali, and D. Knott *et al.*, "The 'big data' revolution in healthcare," *McKinsey Quarterly*, 2013.

[8] R. H. Miller and I. Sim, "Physicians' use of electronic medical records: barriers and solutions," *Health Affairs*, vol. 23, pp. 116–126, 2004.

[9] B. S. Weir, *Genetic Data Analysis*, U.S.: Sinauer Associates, Inc., 1990.

[10] K. Demyttenaere, R. Bruffaerts and J. Posada-Villa *et al.*, "Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys," *JAMA: The Journal of the American Medical Association*, vol. 291, pp. 2581–2590, 2004.

[11] J. Q. Liu, Q. R. Wang, *and* J. F. Wan *et al.*, "Towards real-time indoor localization in wireless sensor networks," in *IEEE 12th International Conf. on Comput. and Inform. Technology*, Cheng du, China, 2012, pp. 877–884. doi: 10.1109/CIT.2012.182.

[12] T. Gao, T. Massey, and L. Selavo *et al.*, "The advanced health and disaster aid network: A light-weight wireless medical system for triage," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 1, no. 3, pp. 203–216, 2007. doi: 10.1109/TBCAS.2007.910901.

[13] A. Georgiades, A. Sherwood,and E. C. Gullette *et al.*, "Effects of exercise and weight loss on mental stress-induced cardiovascular responses in individuals with high blood pressure," *Hypertension*, vol. 36, pp. 171–176, 2000.

[14] K. G. Ong, K. F. Zeng, and C. A. Grimes, "A wireless, passive carbon nanotube-based gas sensor," *IEEE Sensors Journal*, vol. 2, no. 2, pp. 82–88, 2002. doi: 10.1109/JSEN.2002.1000247.

[15] J. Welch, F. Guilak, and S. D. Baker, "A wireless ECG smart sensor for broad application in life threatening event detection," in *26th Annual Int. Conf. IEEE Engineering in Medicine and Biology Society (IEMBS' 04)*, 2004, pp. 3447–3449.

[16] T. Polk, W. Walker, and A. Hande *et al.*, "Wireless telemetry for oxygen saturation measurements," in *IEEE Biomedical Circuits and Systems Conference (BioCAS 2006)*, 2006, pp. 174–177.

[17] S. Sesia, I. Toufik, and M. Baker, *LTE: the UMTS Long Term Evolution*, Wiley Online Library, 2009.

[18] M. Chen, S. W. Mao, and Y. H. Liu, "Big data: a survey," *Mobile Networks and Applications*, vol. 19, no. 2, pp. 1–39, 2014. doi: 10.1007/s11036-013-0489-0.

[19] B. Peng, B. Cui, and X. M. Li, "Implementation issues of a cloud computing platform," *IEEE Data Eng. Bull.*, vol. 32, pp. 59–66, 2009.

[20] J. Dean and S. Ghemawat, "MapReduce: simplified data processing on large clusters," *Communications of the ACM*, vol. 51, pp. 107–113, 2008.

[21] S. M. Domchek, A. Eisen, *and* K. Calzone *et al.*, "Application of breast cancer risk prediction models in clinical practice," *Journal of Clinical Oncology*, vol. 21, no. 4, pp. 593–601, 2003.

[22] P. W. F. Wilson, R. B. D'Agostino, and D. Levy *et al.*, "Prediction of coronary heart disease using risk factor categories," *Circulation*, vol. 97, no. 18, pp. 1837–1847, 1998.

[23] K. D. Mandl and I. S. Kohane, "Healthconnect: clinical grade patient-physician communication," in *Proc. AMIA Symposium*, 1999, p. 849.

[24] P. Bolton, S. Ganon, and D. Aro, "HealthConnect: a trial of an after-hours telephone triage service," *Australian Health Review*, vol. 25, pp. 95–103, 2002.

[25] L. Annemans, E. Spaepen, and M. Gaskin *et al.*, "Gout in the UK and Germany: prevalence, comorbidities and management in general practice 2000–2005," *Annals of the Rheumatic Diseases*, vol. 67, no. 7, pp. 960–966, 2008.

[26] V. G. Vogel, J. P. Costantino, and D. L. Wickerham *et al.*, "Update of the national surgical adjuvant breast and bowel project study of tamoxifen and raloxifene (STAR) P-2 trial: preventing breast cancer," *Cancer Prevention Research*, vol. 3, pp. 696-706, 2010.

Manuscript received: 2014-04-19

Biographies

Jianqi Liu (liujianqi@ieee.org) is a lecturer at Guangdong Mechanical and Electrical College, China. He received his MS degree in computer application technology from Guangdong University of Technology in 2009. He is currently pursuing his PhD degree at GDUT. His research interests include body area networks, embedded systems, internet of vehicle, and cyber-physical systems. He is a member of the IEEE.

Jiafu Wan (jiafuwan_76@163.com) is an associate professor in the School of Electrical Engineering and Automation, JiangXi University of Science and Technology, China. He received his PhD degree in mechatronic engineering from SCUT in 2008. He has authored or co-authored one book and more than 60 scientific papers. His research interests include cyber-physical systems, internet of things, machine-to-machine communication, mobile cloud computing, and embedded systems. He is a CCF senior member and a member of IEEE, IEEE Communications Society, IEEE Control Systems Society, and ACM.

Shenghua He (2001hsh@21cn.com) is an associate professor at Guangdong Mechanical and Electrical College. He received his MS degree in mechatronic engineering from Guangdong University of Technology in 2004. His research interests include management information system, automatic control system, energy management system, etc.

Yanlin Zhang (ylzhxx@qq.com) is a professor at Guangdong Mechanical and Electrical College. He received his MS degree from the Institute of Optics and Electronics, Chinese Academy of Sciences, in 1998. His research interests include mobile networks, enterprise service bus, and energy-saving technologies in large public buildings.

DOI: 10.3969/j. issn. 1673-5188. 2014. 03. 007
http://www.cnki.net/kcms/detail/34.1294.TN.20140723.1827.001.html, published online 23 July, 2014

Reliability of NFV Using COTS Hardware

Li Mo
(CTO Group, ZTE Corporation, RicTX 75080, USA)

Abstract

This paper describes a study on the feasibility of using commercial off -the -shelf (COTS) hardware for telecom equipment. The study outlines the conditions under which COTS hardware can be utilized in a network function virtualization environment. The concept of silent -error probability is introduced to account for software errors and/or undetectable hardware failures, and is included in both the theoretical work and simulations. Silent failures are critical to overall system availability. Site -related issues are created by combined site maintenance and site failure. Site maintenance does not noticeably limit system availability unless there are also site failures. Because the theory becomes extremely involved when site failure is introduced, simulation is used to determine the impact of those facts that constitutes the undesirable features of using COTS hardware.

Keywords

reliability; COTS hardware; NFV

1 Introduction

The use of commercial off-the-shelf (COTS) hardware in architectural frameworks such as IP multimedia subsystem (IMS) and evolved packet core (EPC) has drawn considerable attention in recent years [1]. However, some operators have legitimate concerns about the overall reliability of COTS hardware, including reduced mean time between failures (MTBF), and other undesirable attributes that are unfamiliar in the traditional telecom industry.

In previous reliability studies, such as [2], the focus has only been on hardware failures, which are characterized by the MTBF and mean time-to-repair (MTTR). In this paper, the concept of silent error is introduced to account for failures induced by software and hardware that cannot be detected by the management system.

The analytical expression for silent error is almost impossible to obtain. For hardware-related silent errors, the rate of error depends on the hardware architecture and type of undetectable hardware failure. For software-related silent errors, the software architecture and coding practice need to be investigated. Because there are a variety of hardware and software issues, the analytical expression for silent error is extremely difficult to obtain.

On the other hand, the silent error probability is relatively easy to obtain using observational data. Any error that cannot be attributed to a known cause by the management system can be classified as a silent error. Probability of such an error is the number of errors without known causes divided by the total number of errors observed by the management system. Silent errors affect system availability in different ways depending on the scenario.

In a typical system, a server with certain network functions has another dedicated server as backup. This is a normal master-slave (1+1 redundancy) configuration of the telecom equipment.

The server with the network functions is called the master server, and the dedicated backup is called the slave server. A 1+1 redundancy scheme is different from a 1:1 load-sharing redundancy scheme in that the slave is "dedicated" to backing up the master. In a 1:1 load-sharing redundancy scheme, both servers have network functions and protect each other at the same time.

Here, we assume a single error for ease of discussion. In any protection scheme, system availability is not affected if the slave experiences a silent error and such an error eventually becomes observable from the system's behavior. In this case, another slave is identified, and the master continues with its network functions. Before the slave becomes fully functional, the system is less fault-tolerant.

If the master experiences a silent error, the data transmitted to the slave could be corrupted. In this case, system availability is affected when the error becomes observable in the system's behavior. When a silent error is detected in the master, both the master and slave need time to recover. This recovery time is almost fixed in the network functions virtualisation (NFV) environment and is the COTS MTTR. In this time interval, the network functions are not available, and this is considered downtime in availability calculations.

The MTBF of COTS hardware is shorter than that of typical telecom-grade because COTS hardware has relaxed design criteria. The time needed to repair COTS hardware is not a random variable and is almost fixed in duration. The MTTR of COTS hardware is the mean time required to bring up a server so that it is ready to serve. In traditional telecom hardware, the time to repair is a random variable, and the MTTR is the mean of that random variable. Because manual intervention is normally required in the telecom environment, the MTTR of COTS hardware is usually assumed to be less than that of traditional telecom equipment.

The most obvious difference between COTS hardware and telecom-grade hardware is related to maintenance procedures and practices. With telecom-grade hardware, care is taken to minimize the effect of maintenance on system availability. With COTS hardware, maintenance is often done in a "cowboy" fashion; that is, reset first and ask questions later.

In this study, an analytical solution is proposed for a situation where there are no site- and maintenance-related issues and one or two dedicated backup COTS servers in the NFV environment. An analytical solution is also proposed for a situation where there is no site failure and only one dedicated backup server. In this situation, site maintenance, which is unique to COTS hardware, is addressed. In order to study issues related to site failure, we constructed a simulator and observed system availability with one or two dedicated backup servers. The results show that COTS hardware, with all its undesirable features, still satisfies the telecom requirements under reasonable conditions.

2 Network Reliability

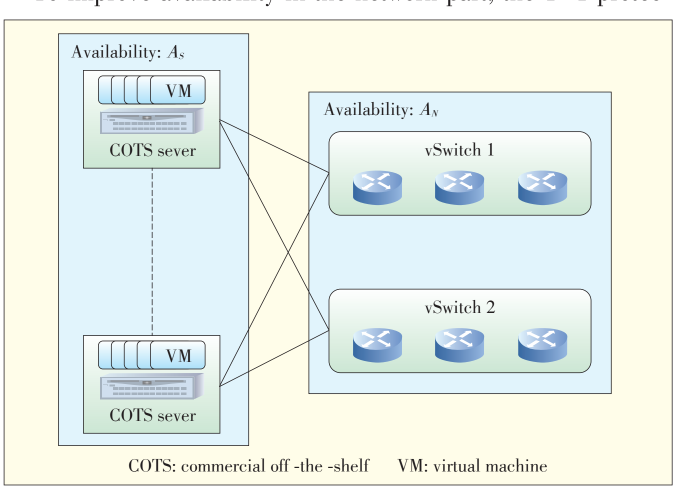
In the NFV environment, reliability of the server part and network part can be analyzed. The server part provides the actual network functions, and the network part connects all the servers with vSwitch (**Fig. 1**).

Overall system availability A_{sys} is a product of the availability of the server part of the system A_s and availability of the network part of system A_N, so that

$$A_{sys} = A_s A_N \quad (1)$$

Given that $A_s < 1$ and $A_N < 1$, then $A_{sys} < A_S$ and $A_{sys} < A_N$. In other words, if five nines are required for system availability, availability in both the server part and network part needs to be better than five nines so that the product of these is more than five nines.

To improve availability in the network part, the 1+1 protection scheme is used (Fig. 1). It is possible for the vSwitch to cover long distance transmission network to connect multiple data centers.



▲**Figure 1. Availability in the network and server parts of the system.**

The mechanisms in the server part that increase availability are not specified. In this study, it is assumed that one or two backup servers support one active server. Normally, if the active server is faulty, one of the backup servers takes over and there is no loss of availability in the server part.

There is a significant difference between the NFV environment and an environment comprising traditional telecom equipment in terms of the time to recover from a server fault. In an environment with traditional telecom equipment, some equipment, e.g., a faulty board, usually needs to be manually changed, and the time needed for restoration after a fault (MTTR) is long.

In an NFV environment, the MTTR is the time required to make another virtual machine (VM) available with the needed software and to re-synchronize the network state data with the server that is currently serving. Hence, the MTTR in the NFV environment is shorter than that of an environment comprising traditional telecom equipment and is a fixed constant.

Also, multiple servers are active in order to share the work load. The failure of individual server only affects a fraction of users, and this has to be taken into account when considering overall system reliability.

Contradictory to common belief, this arrangement neither increases nor decreases overall network availability if the active servers are supported by one or two backup servers. This fact will be elaborated in the later sections from both theoretical point of view and simulations.

3 Availability in the Network Part of the System

Availability in the network part can be analyzed in the traditional way and is affected by 1) the availability of the switch and router that is part of the vSwitch, and 2) the maximum number of hops in the vSwitch. The vSwitch connects the VMs in the NFV environment.

If A_n is the availability of the network element, hops, the availability of a vSwtich with a maximum of h hops is A_n^h. Considering the 1+1 configuration of the vSwtich, A_N is given by

$$A_N = 1 - \left(1 - A_n^h\right)^2 \quad (2)$$

Table 1 shows network availability as a function of the number of hops and per-network availability of network element A_n.

In order to achieve five nines reliability usually demanded by telecom operators, network element reliability needs to be at least four nines if the hop count is greater than 10.

In fact, in order to achieve five nines when per-network element availability is only three nines, the hop count needs to be

▼Table 1. Availability in the network part with various network elements availability and hop count

Network element availability	Hop										
	10	12	14	16	18	20	22	24	26	28	30
0.99	0.990857	0.987092	0.982772	0.977935	0.972614	0.966842	0.96065	0.954066	0.94712	0.939837	0.932244
0.999	0.999901	0.999858	0.999807	0.999748	0.999681	0.999608	0.999526	0.999437	0.999341	0.999237	0.999126
0.9999	0.999999	0.999999	0.999998	0.999997	0.999997	0.999996	0.999995	0.999994	0.999993	0.999992	0.999991
0.99999	1	1	1	1	1	1	1	1	1	1	1

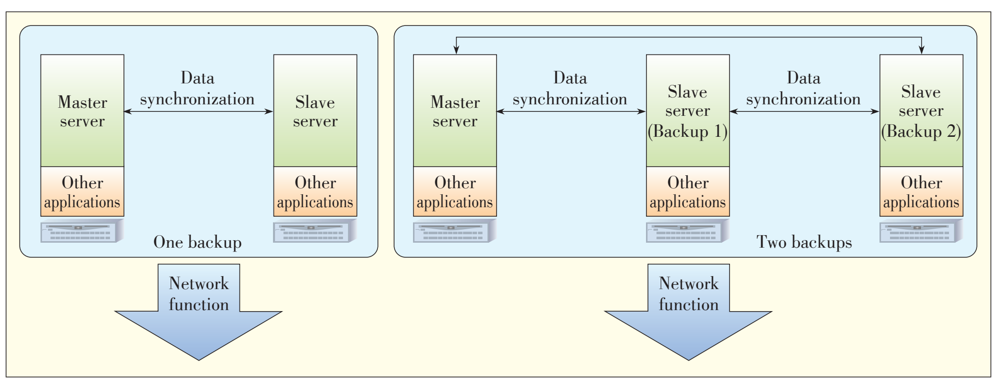
less than two, which is not practical.

4 Redundancy Schemes for Increasing Availability in the Server Part

1+1 redundancy is common in the traditional telecom industry. In the NFV COTS environment, 1+1 redundancy dictates that a set of servers providing network service are protected by other sets of servers. All the servers need to be hosted in different physical machines or even different data centers.

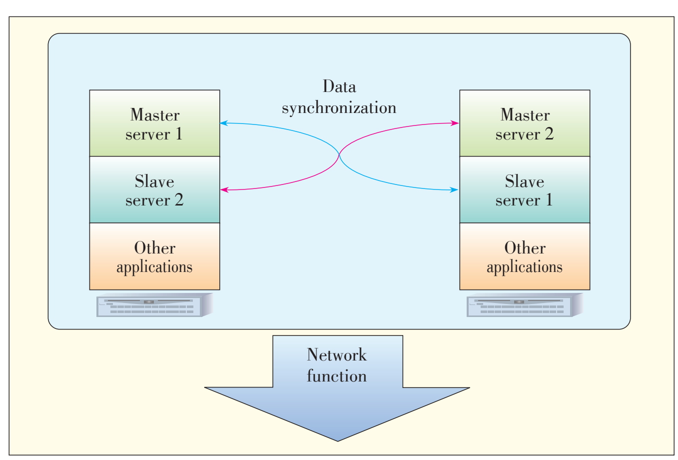
The concept of 1+1 redundancy can also be extended so that multiple servers are used for backup, e.g., two backup servers for one active server. **Fig. 2** shows a setup with one and two dedicated backup servers.

Again, the master and slaves are not co-hosted on the same physical machine. If there is a fault with the master, any slave can become the new master. If there is a fault with a slave, a new VM is designated as the new slave. Before the new slave is ready, the network has reduced protection capabilities. The vSwtich in Fig. 1 provides connectivity between the master and the slaves.



▲Figure 2. 1+1 and 1+2 redundancy.

In traditional telecoms, 1:1 redundancy is commonly used to maximize resource utilization, and servers may share the load. Therefore, a server provides both real-time network functions and protects another server. In 1:1 configuration (**Fig. 3**), masters and slaves are co-hosted. If the physical machine for master 1 and slave 2 is faulty, slave 1 continues the work of master 1, and master 2 loses its protection.



▲Figure 3. 1:1 configuration.

5 Theoretical Analysis of the Server Part and Overall System Availability

In this section, a continuous-time Markov chain [3]–[5] is described for analysis of system availability in the server part.

5.1 Availability for 1+1 Server Configuration

In a practical environment, critical applications, such as the virtualized IP Multimedia Subsystem (vIMS) and virtual Evolved Packet Core (vEPC), are protected from single-server failure. As discussed in section 4, there are a number of schemes to protect against single-server failure. In this section, the 1+1 scheme is discussed, and the continuous-time Markov chain is used to model the 1+1 protected system.

The most significant types of server failure are observable hardware failure; observable non-corrupting software failure, and silent failure.

With observable hardware failure, the faulty component is obvious, and the backup server continues the work of the faulty server without interrupting service. In the worst case scenario, a few services in a transient state, e.g., in the process of establishing a voice session, are affected. Services in a stable state, e.g., an established voice session, are affected.

With an observable, non-corrupting software failure, such as a runaway process, the state of services is not corrupted in the backup server. In this case, the fault recovery is the same as that in the case of an observable hardware failure.

With a silent failure, the cause of the failure is not obvious. The silent failure affects service only if it occurs on the master or current working server. A silent failure affects system availability in the following two ways:

- When the silent failure becomes observable, the network states in both the active server and backup server(s) have been corrupted already, and the whole system needs to be reset.
- When the silent failure becomes observable, the nature of the failure is not obvious, and the system continues to depend on the master or current working server to carry out work because the backup server(s) are assumed to be faulty. Therefore, a reset of the whole system is inevitable.

Fig. 4 shows the Markov state transition for a system with only one backup in a 1+1 configuration.

In Fig. 4, λ is the inverse of the server MTBF, and μ is the inverse of the server MTTR. Both MTBF and MTTR are exponentially distributed random variables. The silent fault probability is denoted s.

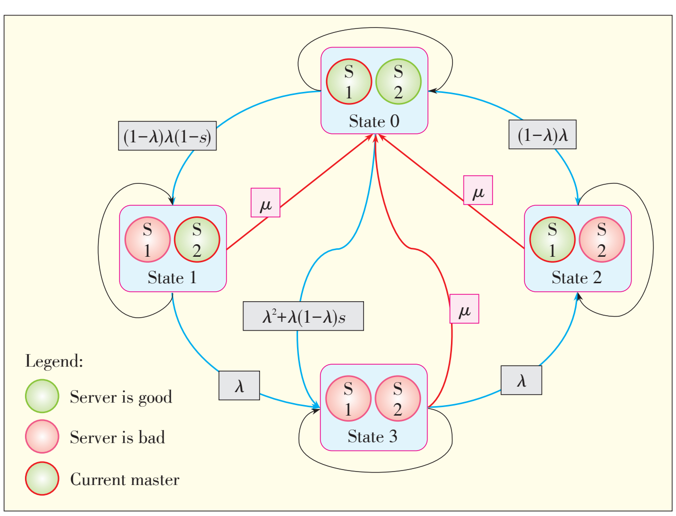
To make the state transition in Fig. 4 valid, the ranges of $\lambda, s, and\, \mu$ are restricted:

$$s \in [0,1],\ \lambda \in (0,1),\ \mu > 0 \quad (3)$$

The real limitation in (3) is that λ has to be less than 1, which means that the individual server does not have MTBF greater than 1 hour.

The probability of being in the states S0, S1, S2, and S4 in the steady state is denoted P_0, P_1, P_2, P_3, respectively. Then, Chapman-Kolmogorov equation for continuous-time Markov chain can be used to express P_3:

$$P_3 = \frac{\lambda}{(\lambda+\mu)(\lambda^2-2\lambda-\mu)}\left(\lambda^2+(s\mu-\mu-2)\lambda-s\mu\right) \quad (4)$$



▲Figure 4. Markov state transition for a system with only one backup in a 1+1 configuration.

$$A_s^{1+1} = 1 - P_3 = 1 - \frac{\lambda^2}{(\lambda+\mu)\left(1-(1-\lambda)^2+\mu\right)}(2+\mu-\lambda) - \frac{\lambda(1-\lambda)\mu}{(\lambda+\mu)\left(1-(1-\lambda)^2+\mu\right)}s \quad (5)$$

Then we have:

$$A_s^{1+1} = \frac{\mu\left(\mu+3\lambda-2\lambda^2\right)}{(\lambda+\mu)\left(1-(1-\lambda)^2+\mu\right)} - \frac{\lambda(1-\lambda)\mu}{(\lambda+\mu)\left(1-(1-\lambda)^2+\mu\right)}s \quad (6)$$

When $s=1$, there is a consistent silent fault on the master side, and the system operates as if there is no protection. Then we have:

$$A_s^{1+1} = \frac{\mu}{\lambda+\mu} = A_s^0 \quad (7)$$

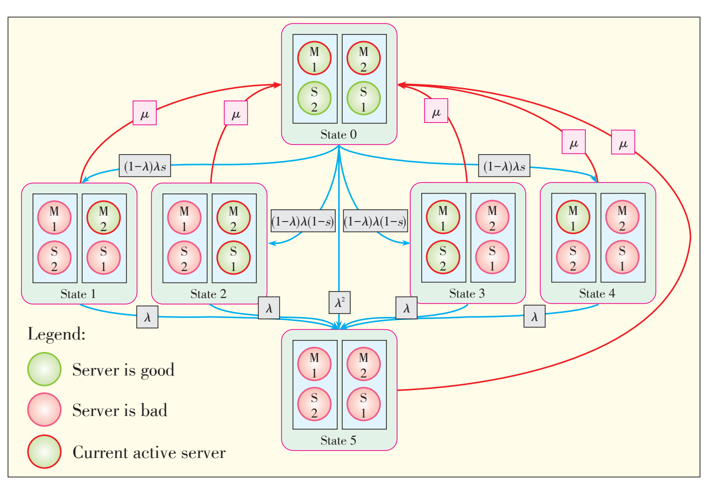
5.2 Availability for 1:1 Server Configuration

The Markov reliability model can also be used in the case of 1:1 load sharing. **Fig. 5** shows the state transition of significance. Trivial state transitions, e.g., S0 to S0 in the case of 1+1 configuration (Fig. 4), are omitted for clarity.

The relationship between the probability of being in any state is given by P_i, $i = 0, 1, 2, 3, 4, 5$. Therefore, we have the following algebraic relationships based on the Global Balance Equations of the continuous-time Markov chain:

$$P_1 = \frac{\mu(1-\lambda)\lambda}{\left(1-(1-\lambda)^2+\mu\right)(\lambda+\mu)}s \quad (8)$$

$$P_4 = \frac{\mu(1-\lambda)\lambda}{\left(1-(1-\lambda)^2+\mu\right)(\lambda+\mu)}s \quad (9)$$



▲Figure 5. Significant Markov state transitions for a system with 1:1 load -sharing.

$$P_5 = \frac{\lambda^2}{\left(1-(1-\lambda)^2+\mu\right)(\lambda+\mu)}(2+\mu-\lambda) \tag{10}$$

The server part of the system is unavailable in S5 and only half available in S1 and S4. Therefore, the availability of a system with 1:1 load sharing is:

$$A_S^{1:1} = 1 - P_5 - \frac{1}{2}P_1 - \frac{1}{2}P_4 \tag{11}$$

$$A_S^{1:1} = 1 - \frac{\lambda^2}{\left(1-(1-\lambda)^2+\mu\right)(\lambda+\mu)}(2+\mu-\lambda) - \frac{\mu(1-\lambda)\lambda}{\left(1-(1-\lambda)^2+\mu\right)(\lambda+\mu)}s \tag{12}$$

Combining (5) and (12), we have:

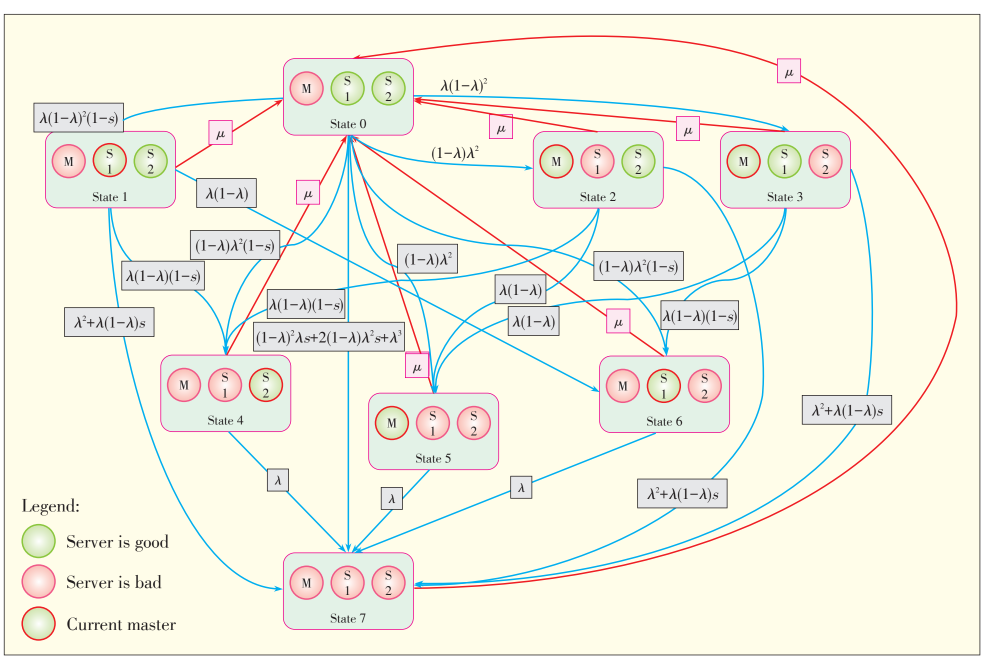
$$A_S^{1:1} = A_S^{1+1} \tag{13}$$

5.3 Availability for a Server Configuration with Dual Backups

Fig. 6 shows the non-trivial Markov state transition for a system with two backups.

The probability at state i is given by P_i, where $i = 0, 1, 2, 3, 4, 5, 6, 7$. As with 1+1 protection, the restrictions on $\lambda, s,$ and μ in (3) are still assumed in this part of study.

Using the Chapman-Kolmogorov equation for continuous-time Markov chains in the steady state, the probability of being in state S0 is denoted P_7 and given by



▲Figure 6. Significant Markov state transition for a system with two backups.

$$P_7 = \frac{\lambda}{(\lambda+\mu)\left(1-(1-\lambda)^2+\mu\right)\left(1-(1-\lambda)^3+\mu\right)}\left(-\lambda^5 + \left(\left(s^2-3s+2\right)\mu+5\right)\lambda^4 + \left(\left(-3s^2+10s-6\right)\mu-9\right)\lambda^3 + \left((1-s)\mu^2+\left(3s^2+6-13s\right)\mu+6\right)\lambda^2 + (6-s)s\mu\lambda + s\mu^2\right) \tag{14}$$

The availability of this configuration is

$$A_S^{1+2} = 1 - P_7 = \frac{\mu}{\left(1-(1-\lambda)^3+\mu\right)(\lambda+\mu)\left(1-(1-\lambda)^2+\mu\right)}\left(\left(-s^2+3s-3\right)\lambda^5 + \left(3s^2-10s+12\right)\lambda^4 + \left(-3s^2+\mu s+13s-19\right)\lambda^3 + \left(s^2-6s-4\mu+11\right)\lambda^2 + (6-s)\mu\lambda + \mu^2\right) \tag{15}$$

The equivalent expression of A_s^{1+2} can also be given as

$$A_S^{1+2} = \frac{\mu}{\mu+\lambda} + \frac{\lambda^2\mu(1-\lambda)^3(1-s)^2}{\left(1-(1-\lambda)^3+\mu\right)(\lambda+\mu)\left(1-(1-\lambda)^2+\mu\right)} + \frac{\lambda\mu(1-\lambda)\left(4\lambda+\mu-3\lambda^2+\lambda^3+\lambda\mu\right)(1-s)}{\left(1-(1-\lambda)^3+\mu\right)(\lambda+\mu)\left(1-(1-\lambda)^2+\mu\right)} \tag{16}$$

When $s = 1$, we have

$$A_S^{1+2} = \frac{\mu}{\mu+\lambda} = A_S^0 \tag{17}$$

When the master is always corrupting the data, system reliability the same as that when there is no protection at all.

6 Site Maintenance

It is important to consider site maintenance when using COTS hardware. To evaluate the effect of site maintenance, the following assumptions are made:

- Site maintenance can only occur if there is no failure in both the working and backup server(s).
- Site recovery, i.e., reverting to the original setup, can only occur if there is no failure in both the working and backup server(s).

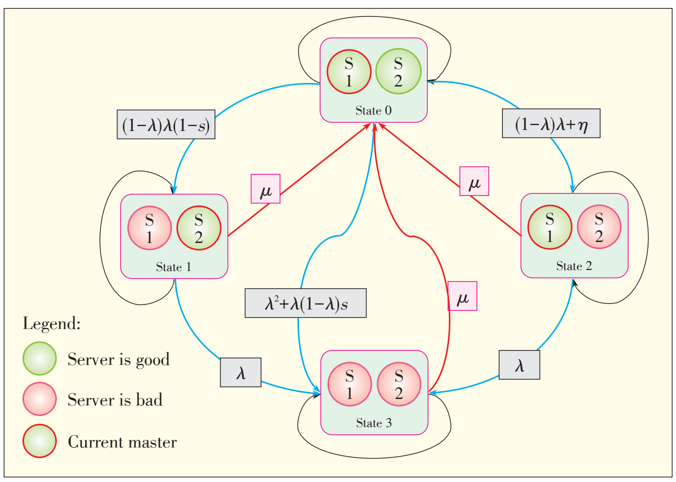
In the following analysis, two distinctive cases are considered for 1+1 configuration:

- Three or more sites of equal capability that cannot be reverted after maintenance. In this case, the third site is used after one site has been put into maintenance mode. During maintenance, the system is still protected by two sites. After maintenance, the system will not revert to its original operation sites.

- Two sites that can be reverted after maintenance. During maintenance, the system has 1+1 protection, but both the mater and slaves operate on the same site. After maintenance, the system reverts to its original mode of operation. This reversion only occurs if both servers are in operation mode.

6.1 Case 1: Non-Revertive After Maintenance with 1+1 Configuration

In the Markov transition diagram for this case (**Fig. 7**), site maintenance probability is assumed to be exponentially distributed, and the inverse mean time-between-maintenance is denoted η. The global balance equation is given by



▲**Figure 7. State transition for non-revertive after maintenance in 1+1 mode.**

$$M=\begin{bmatrix} 1-(1-\lambda)^2+\eta+\mu & 0 & 0 & 0 \\ -\lambda(1-\lambda)(1-s) & \lambda+\mu & 0 & 0 \\ -\lambda(1-\lambda)-\eta & 0 & \lambda+\mu & 0 \\ -\lambda^2-\lambda(1-\lambda)s & -\lambda & -\lambda & \mu \end{bmatrix} \begin{bmatrix} Pn0 \\ Pn1 \\ Pn2 \\ Pn3 \end{bmatrix} = \begin{bmatrix} \mu \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (18)$$

Therefore, the availability is given by

$$A^{1+1}_{MaintN} = 1 - \frac{\lambda\left(\eta+2\lambda-\lambda^2+s\mu+\lambda\mu-s\lambda\mu\right)}{(\lambda+\mu)\left(1-(1-\lambda)^2+\mu+\eta\right)} \quad (19)$$

Given the constraints $\mu>0$, $\eta>0$ and $1>\lambda>0$, we can verify that

$$A^{1+1}_{MaintN} > 1 - P^{1+1}_3 - \frac{\eta\lambda}{\mu^2} = A^{1+1}_s - \frac{\eta\lambda}{\mu^2} \quad (20)$$

A^{1+1}_s is the availability for the 1+1 backup case given by (5).

Equation (20) provides the upper bound of degradation by introducing site maintenance. Given the nominal value of $\mu=10$ and $\eta=$ 1/1000 (i.e., doing maintenance every 1000 hours), degradation is negligible. In the case of dual backup, the effect of site maintenance is expected to be smaller than in the case of 1+1 and is therefore not pursued here.

6.2 Case 2: Revertive After Maintenance with 1+1 Configuration

In the case of two sites, operation needs to be revertive in order to maintain the same level of fault tolerance. **Fig. 8** shows the Markov state transition for 1+1 configuration when the principles for initiating maintenance activity and revertive operations are observed.

In Fig. 8, site maintenance probability is assumed to be exponentially distributed, and the inverse mean time-between-maintenance is denoted η. The recovery time is also assumed to be exponentially distributed, and the inverse MTTR is denoted γ.

The global balance equation is given by

$$M=\begin{bmatrix} 1-(1-\lambda)^2+\eta+\mu & 0 & 0 & 0 & \mu & \mu & \mu & \mu \\ -\lambda(1-\lambda)(1-s) & \lambda+\mu & 0 & 0 & 0 & 0 & 0 & 0 \\ -\lambda(1-\lambda) & 0 & \lambda+\mu & 0 & -\gamma & 0 & 0 & 0 \\ -\lambda^2-\lambda(1-\lambda)s & -\lambda & -\lambda & \mu & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1-(1-\lambda)^2+\gamma & -\mu & -\mu & -\mu \\ 0 & 0 & 0 & 0 & -\lambda(1-\lambda)(1-s) & \lambda+\mu & 0 & 0 \\ -\eta & 0 & 0 & 0 & -\lambda(1-\lambda) & 0 & \lambda+\mu & 0 \\ 0 & 0 & 0 & 0 & -\lambda^2-\lambda(1-\lambda)s & -\lambda & -\lambda & \mu \end{bmatrix} \quad (21)$$

$$M\begin{bmatrix} Ps0 \\ Ps1 \\ Ps2 \\ Ps3 \\ Ps4 \\ Ps5 \\ Ps6 \\ Ps7 \end{bmatrix} = \begin{bmatrix} \mu \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (22)$$

Solving the linear equation (22), we have the site availability:

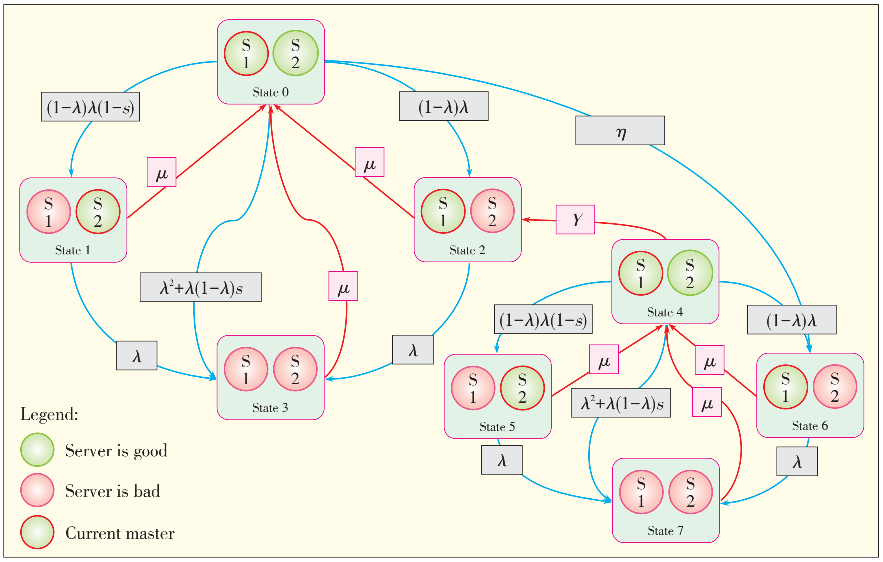
$$A^{1+1}_{siteM} = 1 - Ps3 - Ps7 = 1 - \frac{\gamma\lambda\left(\eta+s\mu-\lambda^2+\lambda(2+\mu-s\mu)\right)}{(\lambda+\mu)\left(2\eta\gamma+(\eta+\gamma)(-\lambda^2+2\lambda+\mu)\right)} - \frac{\eta\lambda\left(\gamma+s\mu-\lambda^2+\lambda(2+\mu-s\mu)\right)}{(\lambda+\mu)\left(2\eta\gamma+(\eta+\gamma)(-\lambda^2+2\lambda+\mu)\right)} \quad (23)$$

Given the constraints $\mu>0$, $\gamma>0$, $\eta>0$, and $1>\lambda>0$, we can verify that

$$A^{1+1}_{siteM} > 1 - P^{1+1}_3 - \frac{2\eta\lambda}{\mu^2} = A^{1+1}_s - \frac{2\eta\lambda}{\mu^2} \quad (24)$$

A^{1+1}_s is the availability for 1+1 backup case given by (5).

Equation (24) provides the upper bound of degradation by introducing site maintenance with revertive operations. Given the nominal value of $\mu=10$ and $\eta=1/1000$ (i.e., doing



▲**Figure 8. Markov state transition with site maintenance and revertive after maintenance operations.**

maintenance every 1000 hours), degradation is negligible. In the case of dual backup, the effect of site maintenance is expected to be smaller and is therefore not pursued here.

Therefore, even with associated site maintenance, COTS hardware does not affect system availability.

7 Simulation of Availability in the Server Part

In the above theoretical analysis of availability in the server part, neither site maintenance (e.g. software upgrade, patching, etc. that affects the whole site) nor site failure (e.g., caused by natural disasters) are taken into account.

Although it is relatively easy to obtain a closed-form solution for system availability in an ideal case without site-related issues, it is extremely difficult to obtain an analytical solution when there are site issues. In this case, we resort to numerical simulation [6] under reasonable assumptions.

7.1 Methodology

A discrete event simulator is constructed to determine the availability in the server part. In the simulator, an active server, i.e., a master that processes network traffic, is supported by a single backup or two backups in another site or in other sites.

The probability of server failure is assumed to follow the bathtub probability distribution (WeiBull distribution). In NFV management, we need to provide servers that are on the flat part of the bathtub distribution. In this case, the familiar exponential distribution can be used.

In the discrete event simulator, each server is scheduled to work for a certain period of time. This is a random variable with exponential distribution, which is commonly used to measure server behavior during the server's useful lifecycle. The mean is given by the MTBF of the server.

The flat part of the bathtub distribution can be related to the normal server MTBF, and the failure density function is given by $F(x) = \frac{1}{MTBF} e^{-\frac{x}{MTBF}}$.

After the period of time that the server is scheduled to work, the server is down for a fixed period of time, i.e., the time needed to start another virtual machine that replaces the one in trouble. This is different for traditional telecom-grade equipment. Here, we assume that there will be another server available to replace the one that goes down. Regardless of the nature of the fault, the down-time for a faulty server is fixed and is the time needed for another server to be ready to take over.

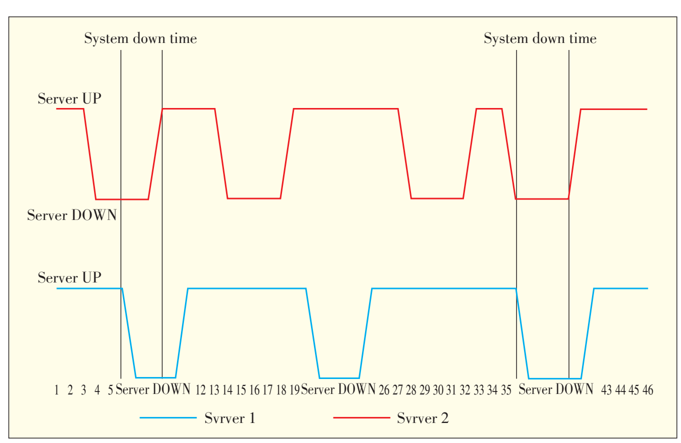
Fig. 9 shows this arrangement for a system with only one backup. Although the server-up time varies, the server down time is of fixed duration.

Servers are hosted at "sites," which are considered to be data centers. In this simulation, during the initial setup, the servers supporting each other were hosted at different sites in order to minimize the effects of site failure and site maintenance.

In order to model system behavior with one or two backups, the concept of protection group is introduced. A protection group comprises a master with one or two slaves at another site or sites (Fig. 2). There may be multiple protection groups inside the network, and each protection group serves a proportion of the users.

A protection group is considered "down" if every server in the group is dead. While the protection group is down, network service is affected, and the network is considered to be down for the group of users that the protection group is responsible for.

The up-time and down-time of the protection group is recorded in the discrete event simulator. Availability in the server



▲**Figure 9. Lifecycle of the servers.**

part is given by

$$Availability_{Server\ Part} = \frac{Weighted\ Protection\ Group\ UP\ Time}{Total\ Elapsed\ Time} \quad (25)$$

where the total time elapsed is the total simulation time elapsed in the discrete event simulator.

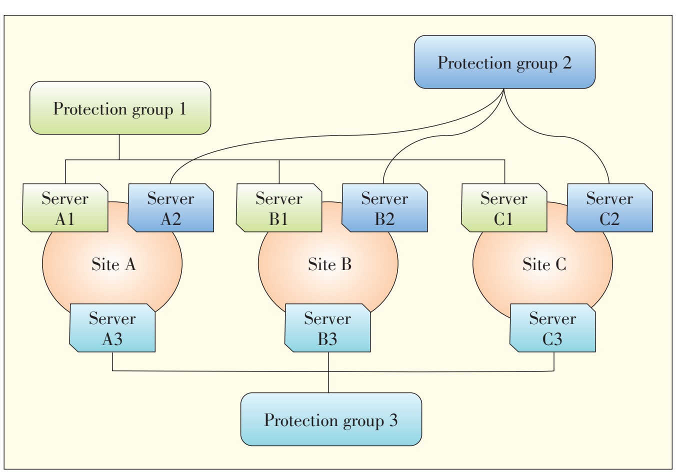
The up-time of the weighted protection group is proportional to the workload of the protection group. In the simulation set-up, the workload is evenly divided between the given number of protection groups.

Fig. 10 shows the protection group, site, and server for a system with two backups. The protection group is an abstract concept, and a proportion of network functions are not available if and only if all the servers in the protection group are not functioning.

Even though the simulator allows each site to have a number of servers (this number is configurable), there is little use for this arrangement. System availability does not change regardless of how many servers per site are used to support the system and as long as there is no change in the number of servers in the protection group. An increase in the number of servers per site is essentially an increase in the number of protection groups. For a long period of time, each protection group will experience similar down-time for the same up-time or will have the same availability.

As in the theoretical analysis, the silent error caused by software or subtle hardware failure only affects the master server. If the master experiences a silent error, both the master and slave have a MTTR, which is the time to incarnate two VMs simultaneously. In this case, this part of the system (or this protection group) is considered to be faulty.

In a reliability study, the focus is on the number of backups for each protection group, and a 1+1 configuration is a typical configuration for one backup mechanism. A load-sharing arrangement such as 1:1 can be viewed as two protection groups. For example, in Fig. 3, master 1 and slave 1 is one protection group, and master 2 and slave 2 is another protection group.



▲**Figure 10. Servers, sites, and protection group.**

From a theoretical point of view, the 1+1 and 1:1 schemes provide similar availability, as in (13). In this work, 1:1 load-sharing is not simulated.

The site also undergoes maintenance. Traditional telecom-grade equipment and COTS hardware differs in terms of maintenance. With telecom-grade equipment, the impact of maintenance on system performance and availability has to be kept to a minimum during the maintenance window. With COTS hardware, maintenance can be more frequent.

To simulate the maintenance aspect of COTS hardware, the simulator puts the site "under maintenance" at a random time. The interval for the site to be working is also assumed to be an exponentially distributed random variable, and the mean is configurable in the simulator. The duration of the maintenance is also a uniform distributed random variable with a configured mean, minimum, and maximum.

In order to put a site under maintenance, there needs to be no faults inside the network, and all servers on the site that are under maintenance are moved to another site. Therefore, no traffic is affected and the probability of site failure is reduced when the site is being maintained.

When a site is back up after maintenance, it attempts to reclaim all its server responsibilities that were transferred due to site maintenance.

If every server is working in a protection group, the protection group rearranges its protection relationship so that each site only has one server in the protection group. The new server on the site that is back up from maintenance has an MTTR to be ready for backup. In this case, there is no loss of service in the system.

If there is at least one working server and one faulty server with a protection group, one working server is added to the protection group. The new server on the site that is back up from maintenance has an MTTR before it is ready for backup. In this case, there is no loss of service in the system.

If no servers are working in the protection group, the protection group gains a new working server from the site that is back up from maintenance. The new server from this site has an MTTR before it is ready for service. The system then provides service when the new server is ready.

A site may also be faulty because of loss of power, thermal issues, or natural disasters. The simulator can also simulate the effects of such events with the site-up duration as an exponentially distributed random variable with configurable mean. The duration of a site failure is expressed as a uniform distributed random variable with configurable mean, minimum, and maximum.

7.2 Simulation Results

In this simulation, both site failure, with MTBF of 10,000 hours, and site maintenance, with mean time between site maintenance of 1000 hours, are considered. The mean time for site failure is assumed to be 12 hours, uniformly distributed between 4 hours and 24 hours. The mean time for site maintenance is 3 hours, uniformly distributed between 4 hours and 48 hours.

The next step is to determine the effect of site failure and maintenance. The very bad site described above has a mean time between site failures that is double the server MTBF, and the mean time between site maintenance is assumed to be 0.1× the server MTBF. **Table 2** shows the availability for a single-backup system.

Availability in the server part is affected by the silent error, and a single redundant piece of hardware provides improvement when the silent error probability is small. **Table 3** shows precise data on availability in the server part with dual backup.

From the tables for single and dual backup, we can see that dual backup is only marginally beneficial when there are site issues. In practice, site issues are inevitable; therefore, a geographically distributed single-backup system is recommended for simplicity.

8 Conclusion

System availability can be divided into two parts: availability in the network and availability in the server. The maximum number of hops determines availability in the network part when the individual network element availability is available. The fault tolerance configuration is assumed to be 1+1. Availability in the server part is mainly determined by

- the availability of an individual server, characterized by its MTBF and MTTR
- silent error probability
- site-related issues, such as maintenances and faults
- protection scheme, i.e., one or two dedicated backups.

The concept of silent error is introduced to account for software errors and errors that cannot be detected by hardware. Availability in the server part is dominated by the silent error if the silent error probability is more than 10%. This is shown in both the theory and simulations. The dual-backup scheme is of marginal benefit, and the added complexity may mean that it is not warranted in the real network.

COTS hardware can provide the same high level of availability as traditional telecom hardware. The undesirable attributes of COTS hardware are modeled into the site-related issues, such as site maintenance and failure. This does not apply to traditional telecom hardware.

Unlike site failure, site maintenance does not noticeably degrade system availability. This applies in both the revertive and non-revertive cases. It is critical for the virtualization infrastructure management to provide as much information about hardware failure as possible in order to increase the availability of the application. In both the theory and simulation, the silent error probability becomes a dominant factor in overall system availability. The silent error probability can be reduced if the virtualization infrastructure management is capable of isolating faults.

▼**Table 2. Availability in the server part with a single backup**

Silent error	Server availability			
	0.990099	0.999001	0.99990001	0.99999
0	0.998971	0.999959	0.9999992	1
0.05	0.998312	0.999908	0.99999464	0.99999949
0.1	0.997918	0.999857	0.99998959	0.99999901
0.15	0.997505	0.999807	0.99998472	0.99999855
0.2	0.996908	0.999771	0.99997957	0.99999804
0.25	0.996599	0.99971	0.99997456	0.9999976
0.3	0.995999	0.999674	0.99996935	0.99999695
0.35	0.995741	0.999625	0.99996403	0.99999651
0.4	0.995177	0.99958	0.99996007	0.99999602
0.45	0.994756	0.999506	0.99995522	0.99999554
0.5	0.994342	0.999466	0.99995014	0.99999502

▼**Table 3. Availability in the server part with dual backup**

Silent error	Server availability			
	0.99009901	0.999001	0.99990001	0.99999
0	0.9999939	0.99999998	1	1
0.05	0.99953558	0.99995032	0.99999511	0.99999948
0.1	0.99910867	0.99990329	0.99999004	0.99999899
0.15	0.99854649	0.99984683	0.99998545	0.99999842
0.2	0.9981346	0.99980209	0.99998048	0.99999792
0.25	0.99763801	0.99975368	0.99997545	0.99999754
0.3	0.99709987	0.99970095	0.99997124	0.99999707
0.35	0.99672872	0.9996602	0.9999657	0.99999655
0.4	0.99615083	0.99960136	0.99996002	0.99999594
0.45	0.99572844	0.9995624	0.99995555	0.99999558
0.5	0.99522474	0.9995184	0.99995225	0.99999503

References

[1] *Network Functions Virtualisation (NFV)*, ETSI GS NFV, 2013.
[2] Applied R \&M Manual for Defense Systems, GR -77, 2012.
[3] A. Leon -Garcia, *Probability, Statistics, and Random Processes for Electrical Engineering*, 3rd Ed. Englewood Cliffs: Prentice Hall, 2008.
[4] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd Ed. New York: McGraw -Hill, 1991.
[5] P. Bremaud, *An Introduction to Probabilistic Modeling*. New York: Springer, 2012.
[6] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd Ed. New York: Cambridge University Press, 1992.

Manuscript received: 2014 -04 -09

Biography

Li Mo (Mo.li@ztetx.com) is the chief architect of the CTO Group, ZTE Corporation. He has more than 20 years' experience in the telecommunications industry. Prior to joining ZTE in 2001, he worked extensively with IBM, Nortel and Fujitsu. His current research interests include SDN, NFV, P2P network, next -generation core networks, IMS, and fixed mobile convergence. He is an active member of the IEEE, ETSI, and ITU -T, where he edited two published recommendations. He has more than 10 approved US patents and many more in progress.

Dr. Mo received his PhD degree in electrical engineering from Queen's University, Canada, in 1989.

DOI: 10.3939/j.issn.1673-5188.2014.03.008
http://www.cnki.net/kcms/detail/34.1294.TN.20140819.0832.001.html, published online 19 August, 2014

Event Normalization Through Dynamic Log Format Detection

Amir Azodi, David Jaeger, Feng Cheng, and Christoph Meinel
(Hasso Plattner Institute, University of Potsdam, 14482 Potsdam, Germany)

Abstract

The analytical and monitoring capabilities of central event repositories, such as log servers and intrusion detection systems, are limited by the amount of structured information extracted from the events they receive. Diverse networks and applications log their events in many different formats, and this makes it difficult to identify the type of logs being received by the central repository. The way events are logged by IT systems is problematic for developers of host-based intrusion-detection systems (specifically, host-based systems), developers of security-information systems, and developers of event-management systems. These problems preclude the development of more accurate, intrusive security solutions that obtain results from data included in the logs being processed. We propose a new method for dynamically normalizing events into a unified super-event that is loosely based on the Common Event Expression standard developed by Mitre Corporation. We explain how our solution can normalize seemingly unrelated events into a single, unified format.

Keywords

event normalization; intrusion detection; event stream processing; knowledge base; security information and event management

1 Introduction

1.1 Event Sources

Events related to security and operation occur in a variety of places in a network. Generally, every computerized device can generate events that can be logged. Both security-information and event-management (SIEM) systems and intrusion-detection (ID) systems are designed to detect unauthorized intrusions into a network. Both of these systems generally operate on the events created by network-enabled devices. Such devices can be classified as hosts or other low-level hardware. On its own, a host can have multiple sources of events at different levels of operation, i.e., at the OS level and application level [1].

The OS is a fundamental source of events on a host. The core of the OS is the OS kernel, which has the highest access privileges and can therefore provide valuable information about the system state and operations performed on the system. This information includes the hardware configuration and state; allocated system resources; security-related operations, such as authentication; and authorization for network and physical access to the host.

Applications running on top of the OS can almost generate endless events relating to many different conditions and are therefore another rich source of event information.

Hardware includes devices that are not host computers in the usual sense, i.e., they run with a static/semi-static software configuration delivered via pre-installed firmware. Devices that fall into this category are network infrastructure devices, such as routers; switches; and peripherals, such as printers or VOIP telephones. Event information from infrastructure devices is especially valuable for SIEM and IDS systems because such devices, especially routers and switches, can intercept connections and control access between nodes.

1.2 Challenges in Event Normalization

When configuring hosts to forward their logs to a log server, an administrator ideally specifies the server's connection details, e.g., the hostname, port, and communication protocol, without configuring the server for individual connections. This means that the client can forward a single event stream containing many diverse events in the system at different levels of the software stack. On the receiving end, the server must separate the different event types and handle them according to their type. The main challenge in handling diverse events lies in identifying the types of events submitted for normalization. Events created by different systems are likely logged using different formats; therefore, there are differences between events that can make normalization difficult.

Each log format has its own limitations. For instance, windows event log favors software debugging and common event expression (CEE) standard allows for limitless arbitrary fields to be added to it. However, CEE does not provide a standardized method of logging even the most basic events related to debugging.

Some log formats, such as Syslog or Snort, are partly or completely unstructured. Unstructured information is a challenge for event normalization because this information cannot be directly mapped to another format if the ontology of the two does not match or overlap for these fields. An example is the message (MSG) field in Syslog.

Even when information in a log format is completely structured, an isolated event often does not provide enough information to be interpreted correctly. Missing information may be details about the log producer, source and target of an activity, or exact time when an event was observed.

1.3 Motivation

When we look at the variety of log formats used by event sources, there is no standard for logging and no agreement on what information is needed to make a log useful for security analysis. SIEM systems, such as Splunk [2], ArcSight [3], Prelude [4] and RSA Envision [5], do not widely use standardized formats or standards focused on alert representation, i.e., events solely focused on system and network security. Considerable effort has been made to address some of these issues [6]–[8]; for example, ArcSight has created its own format, but it is not widely used. We studied SIEM systems and came up with a proposal based on the now-discontinued CEE standardization efforts of Mitre and its partners [9]. We aim to better define the needs of certain sectors of the IT industry and better define the logs they produce so that these logs can be written in a single, all-encompassing format. This format must be highly regulated and meet the needs of network security systems and operational monitoring systems.

1.4 Contribution

The main goal of our research is to provide a solution for an SIEM/IDS system being developed at the Hasso Plattner Institute in Germany. This SIEM/IDS system is actually a real-time event-analysis and monitoring (REAM) system [10] that provides deeper monitoring and analytics for systems across the network. In order to function, the REAM system needs to read arbitrary log files and generate unified events that incorporate all the information gathered from these logs. At the same time, it adds more information where relevant and possible. In this paper, we describe a workable solution for unifying event formats. This solution draws on human knowledge and information embedded within the logs to produce a normalized event-representation model. The solution can be used to correlate events produced by different sources within a network. We describe how named-group regular expressions (NGREs) and a knowledge base can be used to create and populate events with accurate information.

2 Event Normalization

In this paper, the format, syntax, and method of persistence of an event are encompassed within a unified event representation model (UERM). Often, the mapping of elements from one UERM to another is imperfect because there is no logical connection between some elements within the UERMs or an element in one UERM does not exist and is not represented in the other UERM. Therefore, any modern standardized log format has to be highly flexible and have a unified structure. The CEE format provides this flexibility by allowing templates (profiles) to be laid on top of a core profile [11] that comprises elements deemed absolutely necessary for any log entry. These elements include information about the event producer, time the event occurred, and other information. With this layout, software vendors can design their own profiles and lay them on top of the core CEE profile. UERMs such as CEE, CEF, IODEF, IDMEF and Cybox have problems that prevent their use in a REAM system. These problems include lack of openness in the UERM (i.e., new fields can only be added by the vendor), and missing essential features/elements (e.g., lack of support for common security fields, such as CVE) [12]. Such fields are needed to encapsulate other UERMs, such as Syslog, Apache log, and Snort messages. Because CEE is no longer being developed, another UERM, called object log format [13], is being used as the UERM in REAM systems.

In this paper, a security event is the encapsulation of any event related to the confidentiality, integrity, and availability of the system it represents.

2.1 Extracting Event Information from Logs

The steps involved in automatically extracting relevant information in a log line written by people are complex. Issues arise from the unstructured nature of content written and used by people. Some parts of a log entry are loosely structured. Parts such as "client sent HTTP/1.1 request without hostname (see RFC2616 section 14.23): /" can provide important information that, if understood by a machine, could lead to automatic discovery of patterns that are useful in security analysis. One solution implemented in some SIEM systems is collective analysis of all logs of the same type [3], [5]. However, this solution has deficiencies, including omission of important cross-reference information from the analyzed dataset. If a log produced by one app is connected to another log produced by another app, then this connection will be missed. Another issue with creating streams of logs from the same application is that the server has to handle many different connections from a single host and has to be given specific information.

The ultimate goal is to analyze a single, unified, structured set of event information as opposed to multiple sets of information, each of which has its own format. To read logs from the source, understand these logs, and convert them into another format, regular expressions with named capturing groups are necessary. NGREs were first introduced in the Perl programming language, and with Java 1.7, have recently been added to the standard Java API. By using NGREs, it is possible to extract every bit of information from a log line while intelligently assigning them to their mapped attribute within the unified log format. **Fig. 1** shows a complex log line produced by Apache.

Apache access logs is a good example of the easy extraction of information from logs because Apache presents information in a structured manner. With Apache's Combined Log Format specification, it is possible to write a single NGRE that can

```
10.xxx.226.xxx - - [16/Apr/2013:17:05:28 -0400] "GET /static/jquery/
bmi.html?height=200&weight=55  HTTP/1.1" 200 958 "http://small-
tools.com/" "Mozilla/5.0 (Windows NT 6.1; WOW64; rv:20.0) Gecko/
20100101 Firefox/20.0"
```

▲**Figure 1. Apache access log.**

match all possible outputs of one of the most widely used web servers in the world. **Fig. 2** shows an NGRE that can handle almost all access logs generated by the Apache web server. Although it may look complicated, this line only needs to be written once and thereafter can be used by any number of users, none of whom need to know what an NGRE is or how it operates.

After applying the named regular expression from Fig. 2 to the log in Fig. 1, we end up with a list of key-value pairs (**Table 1**).

In most modern programming languages, it is easy to convert such a list into a structured object.

2.2 Creating New Events from a Normalized Event

The results of NGRE matching depend on the event instance

```
(?:(?<network.srcIpv4>(?:[0-9]{1,3}\.){3}[0-9]{1,3})|(?<
  network.srcIpv6>[:\-0-9a-fA-F]+?)|(?<network.srcHost
  >.+?)) - (?:-|(?<user.username>.+)) \[(?<time>.*)\]
  \"(?<application.cmd>(?<application.http.method>[A-Z]+)
  \s(?:(?<application.proto>.*?)://)?(?<network.fqdn
  >[^/]*?)(?:\:(?<network.dstPort>d+))?(?<file.path>/.*?)
  ?(?:\?(?<application.http.queryString>.*?))?(?: HTTP
  /(?<application.http.version>[0-9\.]+)?))\" (?<
  application.http.status>\d+) (?<application.len>\d+)(?:
   "(?:-|(?<application.http.referrer>.*))")?(?:
  "(?:-|(?<application.http.userAgent>.*))")?
```

▲**Figure 2. NGRE for Apache access logs.**

▼**Table 1. Extracted information (key/value)**

network.srcIpv4	181.77.240.127
time	16/Apr/2013:17:05:28 -0400
application.http.method	GET
network.fqdn	-
network.dstPort	80(default value)
file.path	/static/jquery/bmi.html
application.http.queryString	height=175&weight=75
application.http.version	1.1
application.http.status	200
network.ether.len	958
application.http.referrer	http://small-tools.com/
application.http.userAgent	Mozilla/5.0... Firefox/20.0

being matched. The more information inside the event, the more details can be extracted and mapped to their representative values in the new unified log format. In cases where a considerable amount of information is missing from the event instance, human knowledge can greatly help to add the missing information. Therefore, a subsystem was developed to add human knowledge to a unified event. This system (knowledge base) uses a table to store this information alongside the NGREs used to match the event being processed. When a match is made, the information in the knowledge base is used to help build the instance of a unified event. Information extracted from an event using NGREs may be different for two events of the same type. These fields are referred to as dynamic fields. In addition to these fields, there are fields produced by a human and stored in the knowledge base alongside the NGRE. These fields are constant for two or more instances of the same event and are referred to as static fields. To create these fields, system developers have to analyze a representative instance of a given event and include as much information about the instance as possible. This information should be applicable to all instances of the event. Fields such as Time and Producer are not static because they differ from one instance to another.

Different log lines may have the same fields but different values and ranges. To map their value, an interpretation step is necessary. Consider the Priority field. Many log formats specify a field for the event's priority. The more important the producer deems an event, the higher the priority assigned to it. Problems arise when different producers use different formats with different scales or methods for calculating the priority. Some event formats might specify a scale of 1 to 10 whereas others specify a scale of 0 to 255. To map these values directly to the scale used by the UERM representing priority, a script is run on the log line. We implement this script with JavaScript. Short scripts are stored in the database along with the NGRE strings, and the correct script is run when the log format is recognized. The scripts are compiled upon first use, translated into byte code, and run on top of the Java virtual machine (version 8). This provides flexibility because the scripts can be updated or replaced without interrupting the core of the REAMS system.

2.2.1 Defining Event Tags

To help with broader analytical tasks performed on logs, CEE includes a number of tags that can be used by developers to tag an event with a broad, generic label. These tags operate in much the same way as browser bookmark tags or tags used for organizing emails. Logs can be tagged with fields such as Domain, for which one possible value would be Web. This tag can then be used to measure the amount of web-related traffic as a percentage of total network traffic. One simple way to assign these tags to arbitrary logs is to use the knowledge base. Because events are matched, preset tags can be applied to newly created OLF events.

2.2.2 Creating Common Log Entries

Creating a common log format requires a deep understanding and precise mappings of the supported log formats, which are normalized into a single format. One of the more challenging conversions is Syslog to OLF because Syslog has an inherently simple structure but its message (MSG element) has almost limitless complexity.

2.2.3 Event Format Recognition

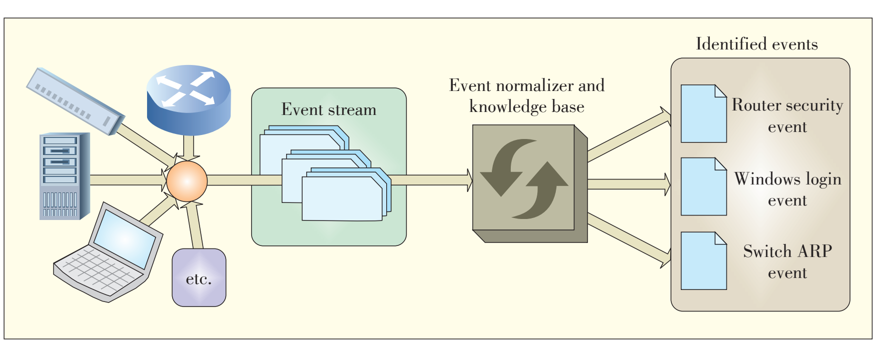
The proposed OLF model comprises hundreds of fields, each mapping to a different aspect of some event. Where possible, duplicates are eliminated and attributes from different logs are mapped to a single field in the OLF. A REAM system uses the approach in [14] to efficiently find the type and variation of the log format currently being processed. Textual metrics and profiling techniques are used to mutate incoming events into an index key, which uses proximity searching to find a known relative of the event instance. When a match is found, an exact NGRE can be used to process the event.

2.2.4 Architecture and Design

The application uses a multithreaded approach: one thread reads the contents of a log file and pushes the logs to a queue, and then multiple worker threads access the queue. Each worker thread attempts to match a regex against the log it has retrieved from the queue using the approach in [14]. When a match is found, information within the log is extracted and inserted into an OLF, which is subsequently persisted to a database. **Fig. 3** shows the steps taken to normalize the events.

3 Related Works

Event normalization, especially incident management, is an ongoing area of research. Much effort has been made in standardizing formats for making event data more persistent, especially for vendors in the IDS domain. The incident object description and exchange format working group (IODEF WG), extended incident handling working group (INCH WG), and MITRE Corporation have provided standards for incident serialization. These standards include the incident object description and exchange format (IODEF) [15] and the intrusion detection message exchange format [8] (IDMEF). The problem with these incident formats is that they are limited to covering alerts only. However, if security breaches have to be analyzed further by inspecting postmortem activities, normally occurring, non-critical activities, such as file access, become more interesting. Looking at log management in general can help overcome the limitations of incident formats. Some approaches to log normalization originated from software products and others originated from standardization institutions. An example of the former is ArcSight's common event format (CEF), proposed in a white paper [6] by Hewlett-Packard. The ArcSight format introduces a flat hierarchy of properties and comprises a set of typical event properties. Two examples limited to web servers are the Common Log Format and Combined Log Format [6]. Both of these were introduced with the Apache web server. Two more generic approaches for event formats are given by MITRE Corporation, which proposed CEE [9] and Cyber Observable eXpression [16] (CybOX). CEE was a promising format because it provided a basic set of common event properties that can further be extended with more event properties as needed. CybOX, on the other hand, is a very complex format that covers most activities in one big format. Nevertheless, CybOX is too bloated to be efficient in a production environment. In addition to research done by standardization bodies, other researchers have investigated ways of efficiently normalizing events. Avourdiadis and Blith [17], [18] propose integrating existing XML-based formats, such as IODEF, IDMEF and Format for INcident information exchange (FINE) [19] into one database. In this database, a core section holds data common in incident messages, and an extensible section can hold additional (uncommon) data in various formats. The mapping of XML elements to database fields is described by an XML document. The limitation of incident messages in XML formats is a big limitation because formats such as Syslog are not fully structured. This makes it hard to map all available information. The authors do not describe how to map such unstructured data.



▲Figure 3. Event detection.

4 Conclusion

In this paper, we have discussed possible improvements to IDS and SIEM systems and the need for better event representation. Two main avenues for regenerating more complete, unified events from their initial representations are: 1) extracting as much information as possible from the event representation being interpreted, and 2) adding human knowledge about the particular event to the unified representation. Most systems attempt to tokenize existing representations of events in order to reproduce them in analytically friendlier formats. However, this is excessively complicated and resource - intensive. Adding support for new log formats often requires considerable effort in terms of development and deployment into existing systems (often via heavy version upgrades). Named-group regular expressions can make the normalization of logs considerably easier. Such expressions can be written to include some logic, e.g., a field may have been omitted from a given instance of a log, and they also allow the application to extract key-value pairs from the log line. The keys can be the fields (or mapped to them) from the UERM used by the system. For the REAM system knowledge base, a UERM was built on the now defunct CEE standard. It is not easy to unify logs, and the task has been attempted by many SIEM system vendors in the past with varying degrees of success. The proposed approach for unifying events decreases the time and effort needed to expand event normalization in an SEIM system and create more complete, intelligent events for analytical purposes.

References

[1] *Guide to computer security log management*, Technology Administration U. S. Department of Commerce, Sept. 2006.
[2] Splunk Inc.. *Splunk enterprise* [Online]. Avaiable: http://www.splunk.com/
[3] Hewlett-Packard. *ArcSight logger* [Online]. http://www8.hp.com/us/en/software-solutions/arcsight-logger-log-management/
[4] CS. *Prelude IDS* [Online]. http://www.prelude-ids.com/en/
[5] EMC2. *RSA envision* [Online]. http://emc.com/security/rsa-envision.htm
[6] H.-P. ArcSight, "Common Event Format," *tech. rep.*, July 2009. Rev. 15.
[7] *The Incident Object Description Exchange Format*, RFC 5070, Dec. 2007.
[8] *The Intrusion Detection, Message Exchange Format (IDMEF)*, RFC 4765, Mar. 2007.
[9] A. Chuvakin, R. Marty, W. Heinbockel, J. Judge, and R. McQuaid, "Common event expression," white paper, CEE Board, June 2008.
[10] *Formerly (SAL)* [Online]. https://hpi.de/meinel/security-tech/network-security/security-analytics.html
[11] T. M. Corporation. *CEE core profile* [Online]. http://cee.mitre.org/language/1.0-beta1/core-profile.html
[12] The MITRE Corporation. *Common vulnerabilities and exposures (CVE)* [Online]. https://cve.mitre.org/
[13] S. Andrey, J. David, A. Amir, G. Marian, C. Feng, and M. Christoph, "Hierarchical object log format for normalisation of security events," in *Proceedings of the 9th International Conference on Information Assurance and Security (IAS2013)*, Kuala Lumpur, Malaysia, 2013, pp. 25-30.
[14] Azodi, "A new approach to building a multi-tier direct access knowledge base for IDS/SIEM systems," *In Proceedings of the 11th IEEE International Conference on Dependable, Autonomic and Secure Computing (DASC2013)*, East Syracuse, NY, USA, 2013.
[15] *The Incident Object Description Exchange Format*, RFC 5070 (Proposed Standard), Dec. 2007.
[16] *The CybOX Language Specification*, the MITRE Corp., Apr. 2012.
[17] N. Avourdiadis and A. Blyth, "Data unification and data fusion of intrusion detection logs in a network centric environment," in *Proceedings of the 4th European Conference on Information Warfare and Security*, Glamorgan, United Kingdom, 2005, pp. 9-20.
[18] N. Avourdiadis and A. Blyth, "Normalising events into incidents using unified intrusion detection-related data," in *Proceedings of the First European Conference on Computer Network Defence School of Computing*, University of Glamorgan, Wales, UK, 2006, pp. 283-296. doi: 10.1007/1-84628-352-3_28.
[19] Y. Demchenko, H. Ohno, R. Danyliw, and G. M. Keeni, "Requirements for format for INcident information exchange (FINE)," Network Working Group, Dec. 2005.

Manuscript received: 2014-04-14

Biographies

Amir Azodi (amir.azodi@hpi.uni-potsdam.de) received his BSc degree in communication networks from Oxford Brookes University. He received his MSc degree in information security from University College London. He is currently a PhD student at the Department of Internet Technologies, Hasso Plattner Institute, Germany. His research interests include event normalization, intrusion detection, attack path detection, and visualization.

David Jaeger (david.jaeger@hpi.uni-potsdam.de) is a PhD student in the IT Security Engineering Team, Hasso Plattner Institute, Germany. From 2006 to 2009, he studied IT systems engineering at the Hasso-Plattner-Institute. He received his BSc degree in 2009 and his MSc degree in 2012. His research interests include intrusion detection, especially attack monitoring and analytics, as well as normalization of security-related information.

Feng Cheng (feng.cheng@hpi.uni-potsdam.de) is a senior researcher heading the IT Security Engineering Team at Hasso Plattner Institute in Germany. His research interests include network security, firewalls, IDS/IPS, security analytics, attack modeling and penetration testing, SOA and Cloud Security. At the Hasso Plattner Institute, he is involved in R&D and teaching activities revolving around new IT security technologies. He has been the principal investigator and project manager for many research projects on IT security, including the project "Physical Separation and its Lock-Keeper Implementation," which was commercialized by Siemens Switzerland (now with Atos Origin) in 2005. He has published more than 30 papers in international conference proceedings and journals. He has been chair, co-chair, coordinator, program committee member, and reviewer for many international workshops and conferences. He received his BEng degree from Beijing University of Aeronautics and Astronautics; he received his MEng degree from Beijing University of Technology; and he received his PhD degree from the University of Potsdam, Germany.

Christoph Meinel (christoph.meinel@hpi.uni-potsdam.de) is scientific director and CEO of the Hasso Plattner Institute, Germany. In 2006, Professor Meinel and Hasso Plattner hosted the 1st National IT Summit of German Chancellor Dr. Angela Merkel at HPI in Potsdam. Dr. Meinel is a member of Acatech (the German National Academy of Science and Engineering) and numerous scientific committees and supervisory boards. Dr. Meinel is a full professor (C4) of computer science and is department chair of internet technologies and systems at the Hasso Plattner Institute. He teaches courses in the Bachelor's degree and Master's degree programs in IT systems engineering and at the HPI School of Design Thinking. He has authored or coauthored nine books and four anthologies and has edited various conference proceedings. He studied mathematics and computer science at Humboldt University of Berlin from 1974 to 1979.

ZTE Communications Guidelines for Authors

• Remit of Journal

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

• Manuscript Preparation

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

• Abstract and Keywords

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

• References

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

• Copyright and Declaration

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

• Content and Structure

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

• Peer Review and Editing

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

• Biographical Information

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

• Acknowledgements and Funding

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

• Address for Submission

magazine@zte.com.cn

12F Kaixuan Building, 329 Jinzhai Rd, Hefei 230061, P. R. China

ZTE COMMUNICATIONS

ZTE Communications has been indexed in the following databases:

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