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

ICACT-TACT JOURNAL

Transactions on Advanced Communications Technology

Volume 5 Issue 3, May. 2016, ISSN: 2288-0003

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

GIRI Global IT Research Institute

Journal Editorial Board

■ Editor-in-Chief

■ Editors

Dr. Youssef SAID, Tunisie Telecom, Tunisia
Dr. Noor Zaman, King Faisal University, Al Ahsa Hofuf, Saudi Arabia
Dr. Srinivas Mantha, SASTRA University, Thanjavur, India
Dr. Shahriar Mohammadi, KNTU University, Iran
Prof. Beonsku An, Hongik University, korea
Dr. Guanbo Zheng, University of Houston, USA
Prof. Sangho Choe, The Catholic University of Korea, korea
Dr. Gyanendra Prasad Joshi, Yeungnam University, korea
Dr. Tae-Gyu Lee, Korea Institue of Industrial Technology(KITECH), korea
Prof. Ilkyeun Ra, University of Colorado Denver, USA
Dr. Yong Sun, Beijing University of Posts and Telecommunications, China
Dr. Yulei Wu, Chinese Academy of Sciences, China
Mr. Anup Thapa, Chosun University, korea
Dr. Vo Nguyen Quoc Bao, Posts and Telecommunications Institute of Technology, Vietnam
Dr. Harish Kumar, Bhagwant institute of technology, India
Dr. Jin REN, North china university of technology, China
Dr. Joseph Kandath, Electronics & Commn Engg, India
Dr. Mohamed M. A. Moustafa, Arab Information Union (AIU), Egypt
Dr. Mostafa Zaman Chowdhury, Kookmin University, Korea
Prof. Francis C.M. Lau, Hong Kong Polytechnic University, Hong Kong
Prof. Ju Bin Song, Kyung Hee University, korea
Prof. KyungHi Chang, Inha University, Korea
Prof. Sherif Welsen Shaker, Kuang-Chi Institute of Advanced Technology, China
Prof. Seung-Hoon Hwang, Dongguk University, Korea
Prof. Dal-Hwan Yoon, Semyung University, korea
Prof. Chongyang ZHANG, Shanghai Jiao Tong University, China
Dr. H K Lau, The Open University of Hong Kong, Honh Kong
Prof. Ying-Ren Chien, Department of Electrical Engineering, National Ilan University, Taiwan
Prof. Mai Yi-Ting, Hsiuping University of Science and Technology, Taiwan
Dr. Sang-Hwan Ryu, Korea Railroad Research Institute, Korea
Dr. Yung-Chien Shih, MediaTek Inc., Taiwan
Dr. Kuan Hoong Poo, Multimedia University, Malaysia
Dr. Michael Leung, CEng MIET SMIEEE, Hong Kong
Dr. Abu sahman Bin mohd Supa'at, Universiti Teknologi Malaysia, Malaysia
Prof. Amit Kumar Garg, Deenbandhu Chhotu Ram University of Science & Technology, India
Dr. Jens Myrup Pedersen, Aalborg University, Denmark
Dr. Augustine Ikechi Ukaegbu, KAIST, Korea
Dr. Jamshid Sangirov, KAIST, Korea
Prof. Ahmed Dooguy KORA, Ecole Sup. Multinationale des Telecommunications, Senegal
Dr. Se-Jin Oh, Korea Astronomy & Space Science Institute, Korea
Dr. Rajendra Prasad Mahajan, RGPV Bhopal, India
Dr. Woo-Jin Byun, ETRI, Korea
Dr. Mohammed M. Kadhum, School of Computing, Goodwin Hall, Queen's University , Canada
Prof. Seong Gon Choi, Chungbuk National University, Korea
Prof. Yao-Chung Chang, National Taitung University, Taiwan
Dr. Abdallah Handoura, Engineering school of Gabes - Tunisia, Tunisia
Dr. Gopal Chandra Manna, BSNL, India

Dr. Il Kwon Cho, National Information Society Agency, Korea
Prof. Jiann-Liang Chen, National Taiwan University of Science and Technology, Taiwan
Prof. Ruay-Shiung Chang, National Dong Hwa University, Taiwan
Dr. Vasaka Visoottiviseth, Mahidol University, Thailand
Prof. Dae-Ki Kang, Dongseo University, Korea
Dr. Yong-Sik Choi, Research Institute, IDLE co., ltd, Korea
Dr. Xuena Peng, Northeastern University, China
Dr. Ming-Shen Jian, National Formosa University, Taiwan
Dr. Soobin Lee, KAIST Institute for IT Convergence, Korea
Prof. Yongpan Liu, Tsinghua University, China
Prof. Chih-Lin HU, National Central University, Taiwan
Prof. Chen-Shie Ho, Oriental Institute of Technology, Taiwan
Dr. Hyoung-Jun Kim, ETRI, Korea
Prof. Bernard Cousin, IRISA/Universite de Rennes 1, France
Prof. Eun-young Lee, Dongduk Woman s University, Korea
Dr. Porkumaran K, NGP institute of technology India, India
Dr. Feng CHENG, Hasso Plattner Institute at University of Potsdam, Germany
Prof. El-Sayed M. El-Alfy, King Fahd University of Petroleum and Minerals, Saudi Arabia
Prof. Lin You, Hangzhou Dianzi Univ, China
Mr. Nicolai Kuntze, Fraunhofer Institute for Secure Information Technology, Germany
Dr. Min-Hong Yun, ETRI, Korea
Dr. Seong Joon Lee, Korea Electrotechnology Research Institute, korea
Dr. Kwihoon Kim, ETRI, Korea
Dr. Jin Woo HONG, Electronics and Telecommunications Research Inst., Korea
Dr. Heeseok Choi, KISTI(Korea Institute of Science and Technology Information), korea
Dr. Somkiat Kitjongthawonkul, Australian Catholic University, St Patrick's Campus, Australia
Dr. Dae Won Kim, ETRI, Korea
Dr. Ho-Jin CHOI, KAIST(Univ), Korea
Dr. Su-Cheng HAW, Multimedia University, Faculty of Information Technology, Malaysia
Dr. Myoung-Jin Kim, Soongsil University, Korea
Dr. Gyu Myoung Lee, Institut Mines-Telecom, Telecom SudParis, France
Dr. Dongkyun Kim, KISTI(Korea Institute of Science and Technology Information), Korea
Prof. Yoonhee Kim, Sookmyung Women s University, Korea
Prof. Li-Der Chou, National Central University, Taiwan
Prof. Young Woong Ko, Hallym University, Korea
Prof. Dimiter G. Velev, UNWE(University of National and World Economy), Bulgaria
Dr. Tadasuke Minagawa, Meiji University, Japan
Prof. Jun-Kyun Choi, KAIST (Univ.), Korea
Dr. Brownson ObaridoaObele, Hyundai Mobis Multimedia R&D Lab , Korea
Prof. Anisha Lal, VIT university, India
Dr. kyeong kang, University of technology sydney, faculty of engineering and IT , Australia
Prof. Chwen-Yea Lin, Tatung Institute of Commerce and Technology, Taiwan
Dr. Ting Peng, Chang'an University, China
Prof. ChaeSoo Kim, Donga University in Korea, Korea
Prof. kirankumar M. joshi, m.s.uni.of baroda, India
Dr. Chin-Feng Lin, National Taiwan Ocean University, Taiwan
Dr. Chang-shin Chung, TTA(Telecommunications Technology Association), Korea

Editor Guide

■ Introduction for Editor or Reviewer

All the editor group members are to be assigned as a evaluator(editor or reviewer) to submitted journal papers at the discretion of the Editor-in-Chief. It will be informed by eMail with a Member Login ID and Password.

Once logined the Website via the Member Login menu in left as a evaluator, you can find out the paper assigned to you. You can evaluate it there. All the results of the evaluation are supposed to be shown in the Author Homepage in the real time manner. You can also enter the Author Homepage assigned to you by the Paper ID and the author's eMail address shown in your Evaluation Webpage. In the Author Homepage, you can communicate each other efficiently under the peer review policy. Please don't miss it!

All the editor group members are supposed to be candidates of a part of the editorial board, depending on their contribution which comes from history of ICACT TACT as an active evaluator. Because the main contribution comes from sincere paper reviewing role.

■ Role of the Editor

The editor's primary responsibilities are to conduct the peer review process, and check the final camera-ready manuscripts for any technical, grammatical or typographical errors.

As a member of the editorial board of the publication, the editor is responsible for ensuring that the publication maintains the highest quality while adhering to the publication policies and procedures of the ICACT TACT(Transactions on the Advanced Communications Technology).

For each paper that the editor-in-chief gets assigned, the Secretariat of ICACT Journal will send the editor an eMail requesting the review process of the paper.

The editor is responsible to make a decision on an "accept", "reject", or "revision" to the Editor-in-Chief via the Evaluation Webpage that can be shown in the Author Homepage also.

■ Deadlines for Regular Review

Editor-in-Chief will assign a evalaution group(a Editor and 2 reviewers) in a week upon receiving a completed Journal paper submission. Evaluators are given 2 weeks to review the paper. Editors are given a week to submit a recommendation to the Editor-in-Chief via the evaluation Webpage, once all or enough of the reviews have come in. In revision case, authors have a maximum of a month to submit their revised manuscripts. The deadlines for the regular review process are as follows:

Evalution Procedure	Deadline
Selection of Evaluation Group	1 week
Review processing	2 weeks
Editor's recommendation	1 week
Final Decision Noticing	1 week

■ Making Decisions on Manuscript

Editor will make a decision on the disposition of the manuscript, based on remarks of the reviewers. The editor's recommendation must be well justified and explained in detail. In cases where the revision is requested, these should be clearly indicated and explained. The editor must then promptly convey this decision to the author. The author may contact the editor if instructions regarding amendments to the manuscript are unclear. All these actions could be done via the evaluation system in this Website. The guidelines of decisions for publication are as follows:

Decision	Description
Accept	An accept decision means that an editor is accepting the paper with no further modifications. The paper will not be seen again by the editor or by the reviewers.
Reject	The manuscript is not suitable for the ICACT TACT publication.
Revision	The paper is conditionally accepted with some requirements. A revision means that the paper should go back to the original reviewers for a second round of reviews. We strongly discourage editors from making a decision based on their own review of the manuscript if a revision had been previously required.

■ Role of the Reviewer

Reviewer Webpage:

Once logined the Member Login menu in left, you can find out papers assigned to you. You can also login the Author Homepage assigned to you with the paper ID and author's eMail address. In there you can communicate each other via a Communication Channel Box.

Quick Review Required:

You are given 2 weeks for the first round of review and 1 week for the second round of review. You must agree that time is so important for the rapidly changing IT technologies and applications trend. Please respect the deadline. Authors undoubtedly appreciate your quick review.

Anonymity:

Do not identify yourself or your organization within the review text.

Review:

Reviewer will perform the paper review based on the main criteria provided below. Please provide detailed public comments for each criterion, also available to the author.

- How this manuscript advances this field of research and/or contributes something new to the literature?
- Relevance of this manuscript to the readers of TACT?
- Is the manuscript technically sound?
- Is the paper clearly written and well organized?
- Are all figures and tables appropriately provided and are their resolution good quality?
- Does the introduction state the objectives of the manuscript encouraging the reader to read on?
- Are the references relevant and complete?

Supply missing references:

Please supply any information that you think will be useful to the author in revision for enhancing quality of the paperor for convincing him/her of the mistakes.

Review Comments:

If you find any already known results related to the manuscript, please give references to earlier papers which contain these or similar results. If the reasoning is incorrect or ambiguous, please indicate specifically where and why. If you would like to suggest that the paper be rewritten, give specific suggestions regarding which parts of the paper should be deleted, added or modified, and please indicate how.

Journal Procedure

Dear Author,

- **You can see all your paper information & progress.**

- **Step 1. Journal Full Paper Submission**

Using the Submit button, submit your journal paper through ICACT Website, then you will get new paper ID of your journal, and send your journal Paper ID to the Secretariat@icact.org for the review and editorial processing. Once you got your Journal paper ID, never submit again! Journal Paper/CRF Template

- **Step 2. Full Paper Review**

Using the evaluation system in the ICACT Website, the editor, reviewer and author can communicate each other for the good quality publication. It may take about 1 month.

- **Step 3. Acceptance Notification**

It officially informs acceptance, revision, or reject of submitted full paper after the full paper review process.

Status	Action
Acceptance	Go to next Step.
Revision	Re-submit Full Paper within 1 month after Revision Notification.
Reject	Drop everything.

- **Step 4. Payment Registration**

So far it's free of charge in case of the journal promotion paper from the registered ICACT conference paper! But you have to regist it, because you need your Journal Paper Registration ID for submission of the final CRF manuscripts in the next step's process. Once you get your Registration ID, send it to Secretariat@icact.org for further process.

- **Step 5. Camera Ready Form (CRF) Manuscripts Submission**

After you have received the confirmation notice from secretariat of ICACT, and then you are allowed to submit the final CRF manuscripts in PDF file form, the full paper and the Copyright Transfer Agreement. Journal Paper Template, Copyright Form Template, BioAbstract Template,

Journal Submission Guide

All the Out-Standing ICACT conference papers have been invited to this "ICACT Transactions on the Advanced Communications Technology" Journal, and also welcome all the authors whose conference paper has been accepted by the ICACT Technical Program Committee, if you could extend new contents at least 30% more than pure content of your conference paper. Journal paper must be followed to ensure full compliance with the IEEE Journal Template Form attached on this page.

➢ How to submit your Journal paper and check the progress?

Step 1. Submit	Using the Submit button, submit your journal paper through ICACT Website, then you will get new paper ID of your journal, and send your journal Paper ID to the Secretariat@icact.org for the review and editorial processing. Once you got your Journal paper ID, never submit again! Using the Update button, you can change any information of journal paper related or upload new full journal paper.
Step 2. Confirm	Secretariat is supposed to confirm all the necessary conditions of your journal paper to make it ready to review. In case of promotion from the conference paper to Journal paper, send us all the .DOC(or Latex) files of your ICACT conference paper and journal paper to evaluate the difference of the pure contents in between at least 30% more to avoid the self replication violation under scrutiny. The pure content does not include any reference list, acknowledgement, Appendix and author biography information.
Step 3. Review	Upon completing the confirmation, it gets started the review process thru the Editor & Reviewer Guideline. Whenever you visit the Author Homepage, you can check the progress status of your paper there from start to end like this, " Confirm OK! -> Gets started the review process -> ...", in the Review Status column. Please don't miss it!

Volume. 5 Issue. 3

Renewable Microgrid State Estimation using the Internet of Things Communication Network

Md Masud Rana, Li Li
Faculty of Engineering and Information Technology
University of Technology Sydney, Broadway, NSW 2007, Australia
Email: 11766084@student.uts.edu.au, mrana928@yahoo.com

***Abstract*—Given the huge concerns all over the world regarding carbon emissions from fossil fuels, energy crisis and global warming, the renewable distributed energy resources (DERs) are going to be integrated in electricity grids, which will make the energy supply more reliable and decrease transmission losses. Regrettably, one of the main practical defies in smart grid planning, control and operation with DERs is the voltage regulation at the distribution field level. This problem motivates the deployment of sensors and actuators in electricity grids so that the voltage regulation can be controlled at the desired level. To do that the measurements from the renewable microgrid state information is transmitted to an energy management center via the internet of things (IoT) based communication network. In other words, the proposed IoT communication infrastructure provides an opportunity to address the voltage regulation challenge by offering the two-way communication links for microgrid state information collection and estimation. Based on this smart grid communication infrastructure, we propose a Kalman filter based state estimation method for voltage regulation of the microgrid. Finally, the effectiveness of the Kalman filter based state estimation method is illustrated using the linear state space model of a microgrid incorporating DERs.**

***Index Terms*—Communication network, distributed energy resource, internet of things, Kalman filter, microgrid, smart grid, state estimation.**

I. Introduction

All over the world, the global warming in one of the major concerns. The key reason behind is the dramatically swelling greenhouse gas emissions from burning fossil fuels and vehicles [1]. In order to diminish this problem, the renewable distributed energy resource (DER) is considered as one of the future electricity generation units [1], [2]. Based on the incentives from governments all over the world, the penetration of DERs is growing promptly. Thus, electricity consumer are participating in the eco-aware global community and the excess amount of energy is sell to the smart grid. Nevertheless, there are significant technical challenges arise in the planning, operation and control of DERs, due to the randomness and weather-dependence in the power generation patterns [3]. Therefore, an unacceptable voltage level may frequently occur at the point common coupling (PCC). This can lead to over-voltage or under-voltage problems for the power network, with undesired voltages appearing at buses of the distribution power network [4], [5]. Driven by this factors, voltage regulators should be installed at planned positions of the distributed feeders [2], [5]. Remarkably, the bidirectional smart grid communication infrastructure between the microgrid and the energy management center can be leveraged to facilitate voltage regulation issues [4]. The key concepts of such intelligent energy management systems are parallel to those of the internet of things (IoT) which can exploit reasonable security and privacy of DERs measurements, seamless interoperability and far-reaching connectivity. To accomplish the goals, the fifth generation (5G) communication network will be the future infrastructure assisting the objectives of the IoT.

A. Related Research

Based on the information and communication network, the smart grid can spread the intelligence from the central control center to the distributed control center, thus enabling accurate state estimation (SE) and wide-area real-time monitoring of the renewable energy sources [6] [7]. First of all, power system SE often practices the weighted least square method that minimizes the sum square of the weighted residuals errors; however, the gain matrix may be ill-conditioned [8]. Later, a comparison between the extended Kalman filter (EKF) and nonparametric belief propagation (NBP) has been implemented for distributed dynamic state estimation [9]. More specially, a NBP method to compute the power system state is developed, showing that the performance of the NBP is better than that of EKF algorithm. In [10], [11], a factor graph based message-passing algorithm for power system state estimation is proposed. Generally, the factor graph entails of variable and factor nodes. The factor nodes are the logical representation of the sensor measurement, whereas the variable nodes do not exists actually [10]. The information can be processed and passed between the variable and factor nodes with definite sum product rules [10], [11]. A BP algorithm has interesting structural properties corresponding to nonlinear feedback dynamical systems in the context of decoding the received signal [12]. Generally, the BP based statistical estimation techniques can provide a better performance if there

Manuscript received Sept. 29, 2015. Part of this work is published in the 12th International Conference on Information Technology-New Generations. This is an invited paper from the ICACT-TACT Journal.

Md Masud Rana Faculty of Engineering and Information Technology, University of Technology, Sydney, NSW 2007, Australia (Corresponding author to provide phone: +61470352998; e-mail: 11766084@student.uts.edu.au)

Li Li are with Faculty of Engineering and Information Technology, University of Technology, Sydney, NSW 2007, Australia.

is no cycle in the graph [13]. This method can converge to the actual system states in the tree like configuration. When cycles are present in the graph, the technique may cause fluctuation and the estimated state may diverge from the actual state [13], [14].Furthermore, the ensemble KF approach uses the probability distribution function of the system state and the data likelihood [15]. Due to the use of stochastic measurement rather than of the whole available data set, it is computationally faster and performs satisfactorily for highly stochastic systems.

B. Key Contributions

This paper proposes an approach for microgrid state estimation using the IoT networks. First of all, a renewable microgrid incorporating multiple DERs is modelled as a continuous linear state space model. This model is transformed to the discrete linear state space system considering the uncertainty. Then the smart sensors are positioned around the microgrid to obtain the measurements. Afterward, the measurements from the microgrid is transmitted to an energy management center via the IoT based 5G communication network. This IoT communication technology affords an opportunity to address the estimation challenge by offering the two-way communication links for microgrid state information collection and estimation. Based on this smart grid communication infrastructure, we propose a KF algorithm for state estimation. The effectiveness of the KF method is verified by numerical simulations using a microgrid incorporating DERs.

The remaining of this manuscript is organized as follows. The fundamental description of the IoT and its vision is described in Section II. Section III explores the IEEE 4-bus distribution system with microgrid model and IoT communication network. In addition, the proposed KF based dynamic SE scheme is described in Section IV, followed by the simulation results and discussions in Section V. This is followed by the conclusion in Section VI.

Notation: Bold face lower and upper case letters are used to represent vectors and matrices, respectively; and $\mathbf{I}$ is the identity matrix.

II. Architecture and Vision of the IoT

The IoT is a vision that encompasses and surmounts several technologies at the confluence of power systems, information technology, medicines, nanotechnology and biotechnology [16], [17]. In fact, the application scenarios of the IoT in diverse areas is illustrated in Fig. 1. The IoT has been considered as the latest revolution in the digital technology after the invention of computers and the internet [16], [18]. From the aspect of electricity network, it brings major benefits to the smart grid infrastructure design. Technically, it represents a world-wide network of heterogeneous things such as smart devices, smart objects, smart sensors, smart actuators, radio frequency identification (RFID) tags and readers, global positioning systems (GPS) and embedded computers [18]. Such things can be deployed and exploited in different physical environments to support diversified cyber physical applications such as information collection, information processing, identification, control and actuation [18], [19]. For clarify of

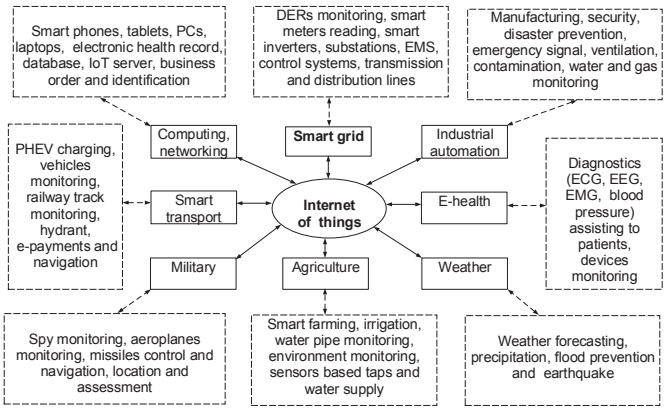


Fig. 1. The application scenarios of the IoT [16].

understanding, Fig. 2 shows the information flow between the cyber and physical space using the IoT infrastructure. It can

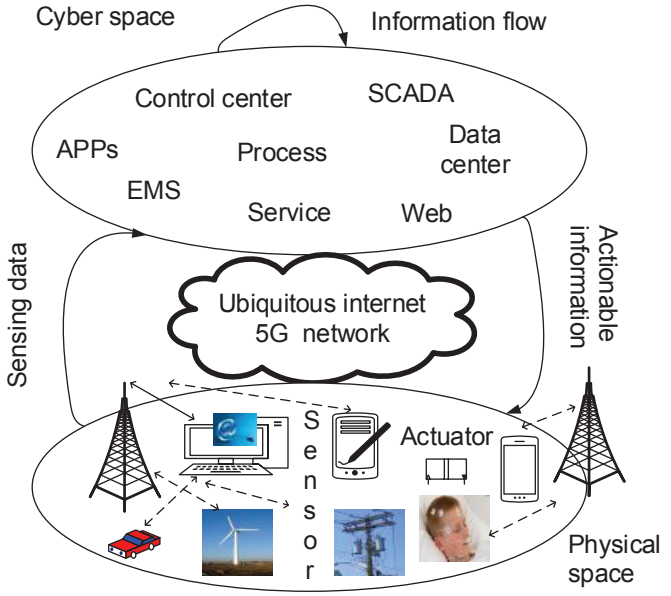


Fig. 2. Information flow between the cyber and physical space using the IoT with 5G networks [18].

be seen that the information produced in the physical space is transmitted to the cyber space for interpretation, which in turn affects the physical environment such as plug in hybrid electric vehicle and smart grid communications [18].

Due to the economic, environmental as well as technical reasons, the energy sector has a growing awareness of smart grid technologies to enhance the efficiency and reliability of electricity networks [1], [20]. From this perspective, renewable DERs such as solar cells, photovoltaic arrays and wind turbines, have been integrated into the grid in the form of smart distribution grids. From the aspect of smart devices and smart metering, they play a vital role for remote monitoring and power systems' state estimation [21]. The reliable state estimation is a key technique to fulfil the automation of power grids. In order to monitor the DER state, the proposed IoT

based communication network architecture for sensing the DER states describes in the next section.

III. DERs Model and IoT Communication Network

This section illustrates the multiple DERs model that is connected to the IEEE 4-bus distribution system, observation model, uniform quantization and IoT communication systems.

A. DERs Connected to the IEEE 4-bus distribution system

Fig. 3 shows IEEE 4-bus distribution test feeders that are interfaced to the local load through converter [22]. We adopt

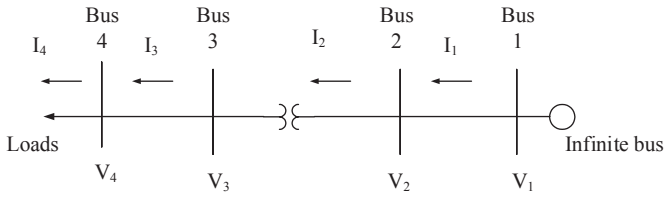


Fig. 3. An illustration of the IEEE 4-bus distribution system.

the model of interconnected DERs from [23], [22], as shown in Fig 4. It is assumed that four DERs are modelled as

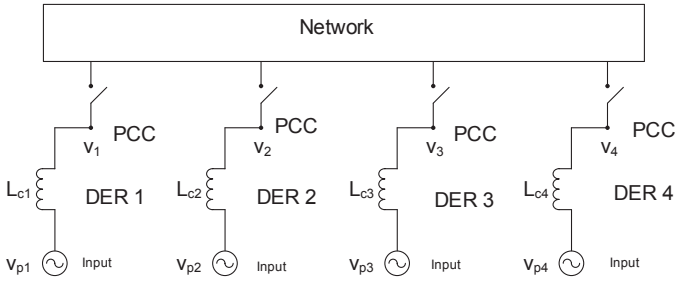


Fig. 4. Four DERs are connected to the power network [22].

voltage sources whose input voltages are denoted by $\mathbf{v}_p = (v_{p1}\ v_{p2}\ v_{p3}\ v_{p4})^T$, where v_{pi} is the i-th DER input voltage. The four DERs are connected to the main power network at the corresponding Point of Common Coupling (PCCs) whose voltages are denoted by $\mathbf{v}_s = (v_1\ v_2\ v_3\ v_4)^T$, where v_i is the i-th PCC voltage. In order to maintain the proper operation of DERs, these PCC voltages need to be kept at their reference values. A coupling inductor exists between each DER and the rest of the electricity network. Now applying the Laplace transformation in this microgrid to obtain the nodal voltage equations. The nodal voltage equation is given by:

$$\mathbf{Y}(s)\mathbf{v}_s(s) = \frac{1}{s}\mathbf{L}_c^{-1}\mathbf{v}_p(s), \tag{1}$$

where $\mathbf{L}_c = diag(L_{c_1},\ L_{c_2},\ L_{c_3},\ L_{c_4})$ and $\mathbf{Y}(s)$ is the admittance matrix of the power network. Based on the typical assumptions of the IEEE 4-bus distribution feeder [22], the admittance matrix is given in (2). Now we can transform the Laplacian form into the linear state space model. The brief conversion can be found in [22]. Normally, the dynamic of the physical state space system is given by:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{n}(t), \tag{3}$$

where $\mathbf{x}(t) = \mathbf{v}_s - \mathbf{v}_{ref}$ is the PCC state voltage deviation, $\mathbf{v}_{ref}$ is the PCC reference voltage, $\mathbf{u}(t) = \mathbf{v}_p - \mathbf{v}_{pref}$ is the DER control input deviation, $\mathbf{v}_{pref}$ is the reference control effort, $\mathbf{n}(t)$ is the zero mean process noise whose covariance matrix is $\mathbf{Q}_n$, the state matrix $\mathbf{A}$ and input matrix $\mathbf{B}$ are given by:

$$\mathbf{A} = \begin{bmatrix} 175.9 & 176.8 & 511 & 103.6 \\ -350 & 0 & 0 & 0 \\ -544.2 & -474.8 & -408.8 & -828.8 \\ -119.7 & -554.6 & -968.8 & -1077.5 \end{bmatrix}, \tag{4}$$

$$\mathbf{B} = \begin{bmatrix} 0.8 & 334.2 & 525.1 & -103.6 \\ -350 & 0 & 0 & 0 \\ -69.3 & -66.1 & -420.1 & -828.8 \\ -434.9 & -414.2 & -108.7 & -1077.5 \end{bmatrix}. \tag{5}$$

The continuous state space model (3) can be written into the following discrete state space form:

$$\mathbf{x}(k+1) = \mathbf{A}_d\mathbf{x}(k) + \mathbf{B}_d\mathbf{u}(k) + \mathbf{n}_d(k), \tag{6}$$

where $\mathbf{A}_d = \mathbf{I} + \mathbf{A}\Delta t$, Δt is the discretization step value, $\mathbf{B}_d = \mathbf{B}\Delta t$ and $\mathbf{n}_d(k) = \Delta t\mathbf{n}(k)$ with the variance $\mathbf{Q}_{nd}$ [24], [25]. The smart sensors can sense the microgrid states to form an observer model as follows:

$$\mathbf{y}(k) = \mathbf{C}\mathbf{x}(k) + \mathbf{w}(k), \tag{7}$$

where $\mathbf{y}(k)$ is the measurement, $\mathbf{C}$ is the measurement matrix and $\mathbf{w}(k)$ is the measurement noise whose variance is $\mathbf{Q}_{wd}$. The observation noise comes from the distributed wireless sensors measurements. The observation information by the wireless sensor networks (WSN) powered by 5G technologies is transmitted to the nearby base station (BS) as shown in Fig. 5. The uniform quantizer of this base station maps each measurement signals to a sequence of bits.

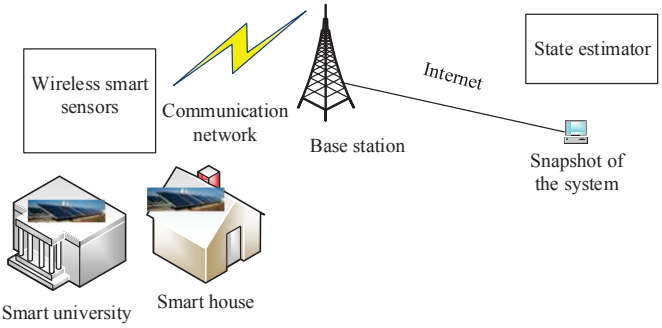


Fig. 5. Interfacing microgrid measurements to the base station.

B. IoT Communication Network

For transmitting the microgrid state information to the energy management system, we used binary phase shift keying (BPSK) as a modulation technique. The bit sequence $\mathbf{b}(k)$ is goes through a BPSK and get modulated signal $\mathbf{s}(k)$. The $\mathbf{s}(k)$ goes through the internet and add noise. To demonstration, Fig. 6 shows the IoT communication network and dynamic state estimation process. The received signal at the energy

$$\mathbf{Y}(s)=(\mathbf{L}_c s)^{-1}+\begin{bmatrix} \frac{1}{0.1750+0.0005s} & \frac{-1}{0.1750+0.0005s} & 0 & 0 \\ \frac{-1}{0.1750+0.0005s} & \frac{1}{0.1750+0.0005s}+\frac{1}{0.1667+0.0004s} & \frac{-1}{0.1667+0.0004s} & 0 \\ 0 & \frac{-1}{0.1667+0.0004s} & \frac{1}{0.1667+0.0004s}+\frac{1}{0.2187+0.0006s} & \frac{-1}{0.2187+0.0006s} \\ 0 & 0 & \frac{-1}{0.2187+0.0006s} & \frac{1}{0.2187+0.0006s}+\frac{1}{12.3413+0.0148s} \end{bmatrix}. \quad (2)$$

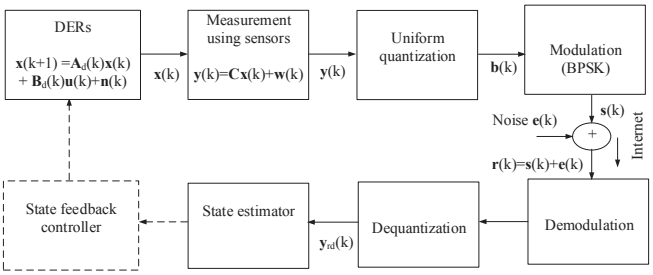


Fig. 6. The IoT communication network and microgrid dynamic state estimation process.

management system is given by

$$\mathbf{r}(k)=\mathbf{s}(k)+\mathbf{e}(k), \quad (8)$$

where $\mathbf{e}(k)$ is additive white Gaussian noise AWGN noise. Then the received sequence is followed by dequantization and finally KF algorithm is used for this microgrid.

IV. Kalman filter based microgrid state estimation method

This section tries to answer the following question: (i) What is the optimal smart grid SE method for the microgrid incorporating multiple DERs?

The discrete time KF is a set of recursive mathematical equations that provides an efficient recursive means to estimate the state of a process in a way that minimizes the mean squared error between the measurement and prediction. The KF operates recursively on streams of the noisy measurers to produce a statistically optimal estimate of the underlying microgrid system states. This method works in two-steps (prediction and correction step). The energy management system computes the following steps [26]:

$$\hat{\mathbf{x}}^{-}(k)=\mathbf{A}_d\hat{\mathbf{x}}(k-1)+\mathbf{B}_d\hat{\mathbf{u}}(k-1), \quad (9)$$

where $\hat{\mathbf{x}}^{-}(k)$ is the microgrid estimate states of the earlier step. The predicted estimate covariance matrix is given by:

$$\mathbf{P}^{-}(k)=\mathbf{A}_d\mathbf{P}^{-}(k-1)\mathbf{A}_d^T+\mathbf{Q}_{nd}, \quad (10)$$

where $\mathbf{P}^{-}(t)$ is the microgrid estimate covariance matrix of the earlier step. The microgrid estimated update state (correction step) is given by

$$\hat{\mathbf{x}}(k)=\hat{\mathbf{x}}^{-}(k)+\mathbf{K}(k)[\mathbf{y}_{rd}(k)-\mathbf{C}\hat{\mathbf{x}}^{-}(k)], \quad (11)$$

where $\mathbf{y}_{rd}(k)$ is the dequantized and demodulated output bit sequences and the Kalman gain $\mathbf{K}(t)$ is given by:

$$\mathbf{K}(k)=\mathbf{P}^{-}(k)\mathbf{C}^T(\mathbf{C}\mathbf{P}^{-}(k)\mathbf{C}^T+\mathbf{Q}_{wd})^{-1}, \quad (12)$$

and

$$\mathbf{P}(k)=\mathbf{P}^{-}(k)-\mathbf{K}(k)\mathbf{C}\mathbf{P}^{-}(k). \quad (13)$$

Based on the KF steps, the energy management system can obtain the predicted distribution of the measurement which is a Gaussian distributed with the expectation given by Eq. (9) and the covariance matrix given by Eq. (10). Considering the aforementioned expressions taking into account, the flow chart for the KF state estimation algorithm is sketched in Fig. 7.

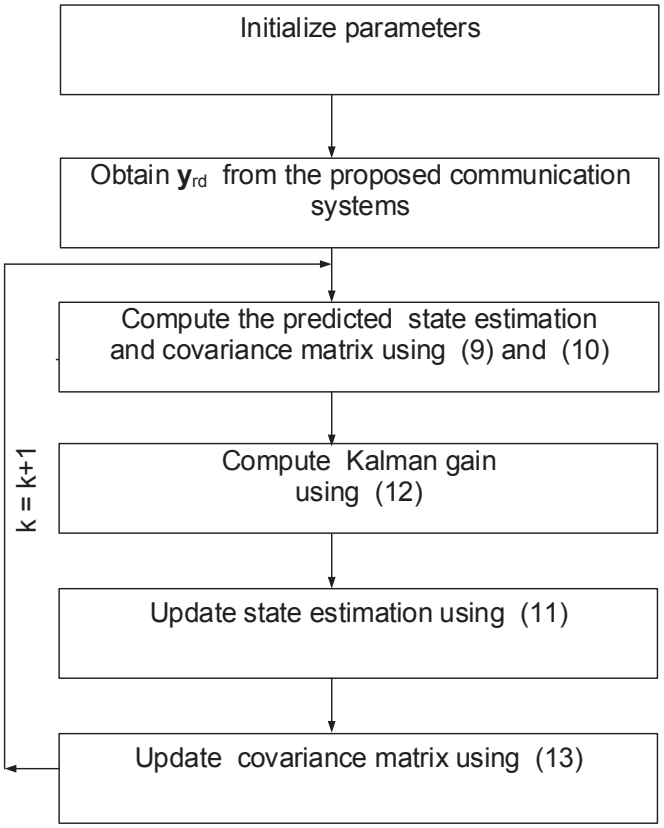


Fig. 7. The flow chart for the Kalman filter based microgrid state estimation.

V. Performance Evaluations and Discussions

We consider four DERs in which the system state is a four-dimensional vector. Each DER is connected to the IEEE 4-bus distributed systems operated as an island mode [27]. The continuous state space system has been approximated to the discrete time state space system with a small step size parameter. The simulation parameters of the IoT networks are summarized in Table I. The simulation of this proposed KF

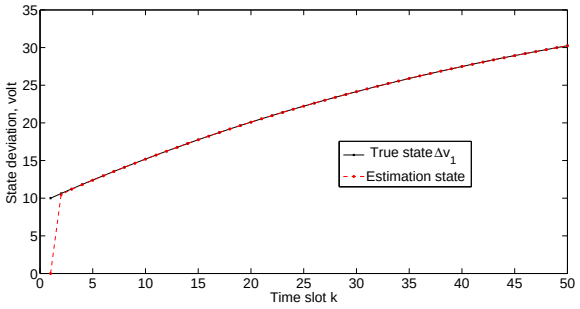


Fig. 8. Δv_1 comparison between the true and estimated state using 4 sensors.

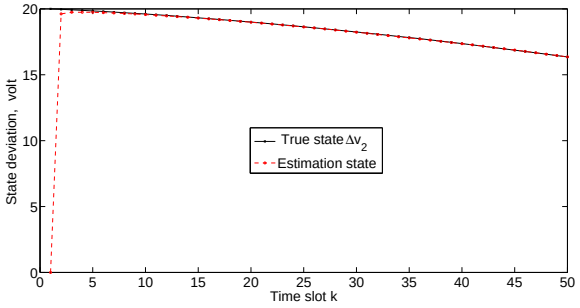


Fig. 9. Δv_2 comparison between the true and estimated state using 4 sensors.

based microgrid SE for the IoT communication network is carried out for two different sensing scenarios.

TABLE I
THE SYSTEM PARAMETERS AND ASSUMPTIONS USING MATLAB.

Parameters	Values	Parameters	Values
Step size Δt	0.00001	Quantization	Uniform
Modulation	BPSK	Channel	AWGN
Time slots	50	Quantization	Uniform 16 bits

A. *Number of sensors equal to states*

First of all, it assumes that the four voltage sensors can sense the four PCC voltage states directly. From the simulation results as shown in Figs. 8 to 11, it can be seen that the proposed KF is able to estimate the PCC state voltages properly and it needs few iterations to track the original states. However, in practical scenario there are the possibility that the smart sensors battery gets low so that they cannot sense the system state properly.

B. *Number of sensors less than states*

When the smart sensors battery gets low and it cannot sense the system state properly. In this case, one assumes that the two voltage sensors are out of order among four sensors to sense the four PCC voltage states. From the simulation results as shown in Figs. 12 to 15, it can be seen that the proposed KF is able to estimate system states properly and it needs more iterations to track the original states. Due the two sensors sensing problems, the PCC state voltages v_2 and v_3 are not

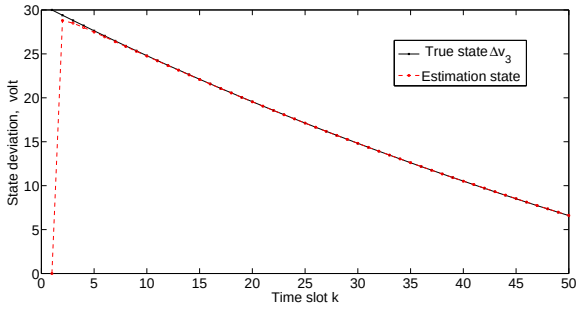


Fig. 10. Δv_3 comparison between the true and estimated state using 4 sensors.

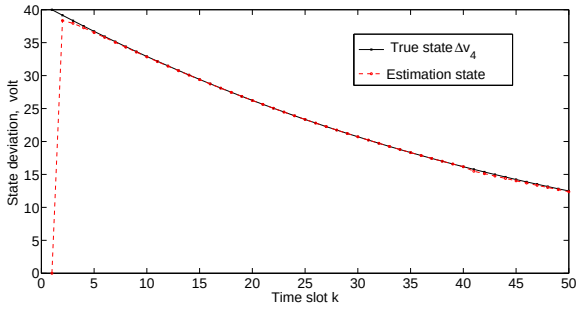


Fig. 11. Δv_4 comparison between the true and estimated state using 4 sensors.

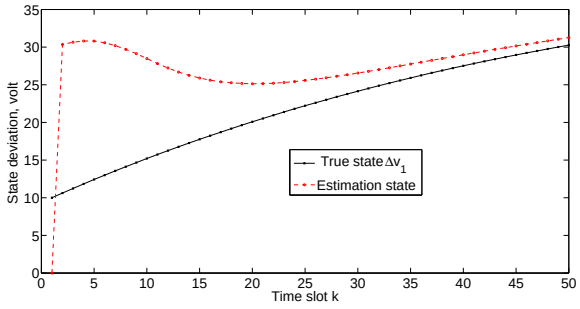


Fig. 12. Δv_1 comparison between the true and estimated state using 2 sensors.

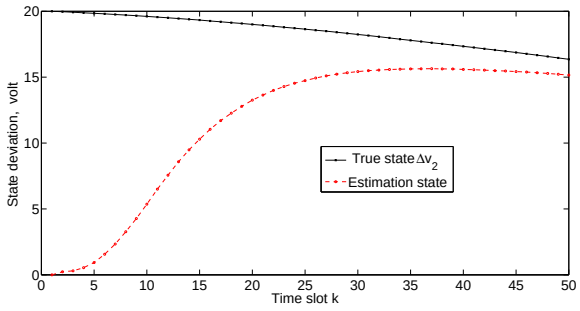


Fig. 13. Δv_2 comparison between the true and estimated state using 2 sensors.

able to be sensed directly by the corresponding sensors. But the KF is able to track these states with small considerable errors. However, there are the possibility that the number of

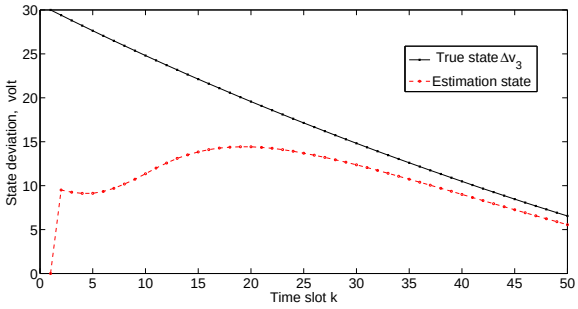


Fig. 14. Δv_3 comparison between the true and estimated state using 2 sensors.

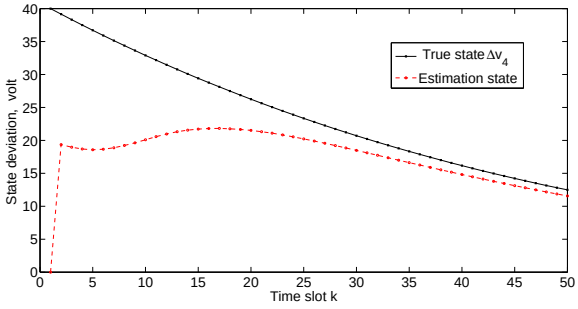


Fig. 15. Δv_4 comparison between the true and estimated state using 2 sensors.

smart sensors is greater than the system states.

C. *Number of sensors greater than states*

Finally, we assume that the number of sensors is greater than the system states. This simulation it is assumed that the observation matrix has eight sensors. From the simulation results as shown in Fig. 16 to 19, it can be seen that the proposed KF is able to estimate system states properly and it needs few iterations to track the original states. Due the more sensors, the PCC voltage are not able to sense by the the sensors correctly. But using the proposed KF is able to track these states with very small errors. Therefore, it is better to use same number of sensors and states to properly estimate system states in the IoT communication network.

VI. Conclusions

This paper addresses the voltage regulation issue from the communication perspective. To do so, wireless sensor network components such as sensors and actuators have been applied into the microgrid to coordinate DER states regulation. In order to transmit the sensing information to the observer, the proposed innovative communication systems have been utilized. Based on this infrastructure, this paper proposes a KF algorithm for centralized DER state estimation. Finally, the effectiveness of the developed approaches is verified by numerical simulations. In the future, we will use least square based Kalman filter in order to obtain the better initial state value.

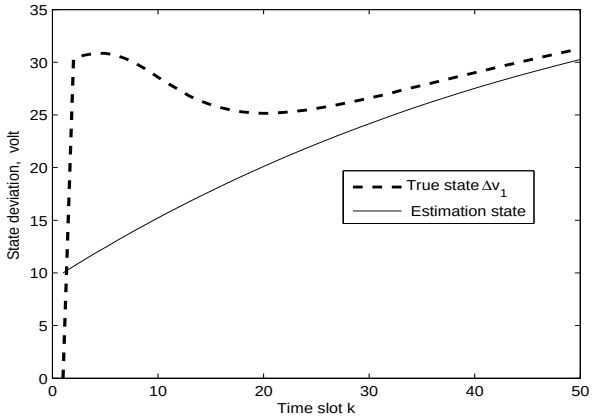


Fig. 16. Δv_1 comparison between the true and estimated state using 8 sensors.

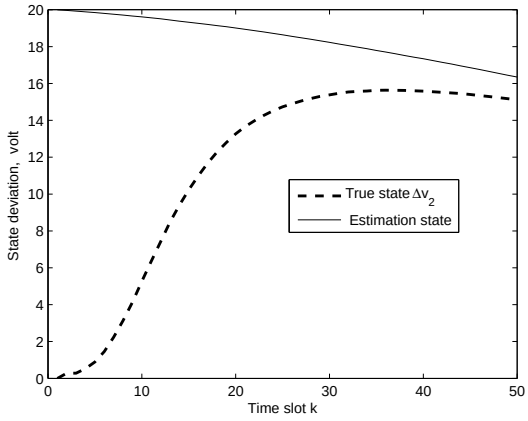


Fig. 17. Δv_2 comparison between the true and estimated state using 2 sensors.

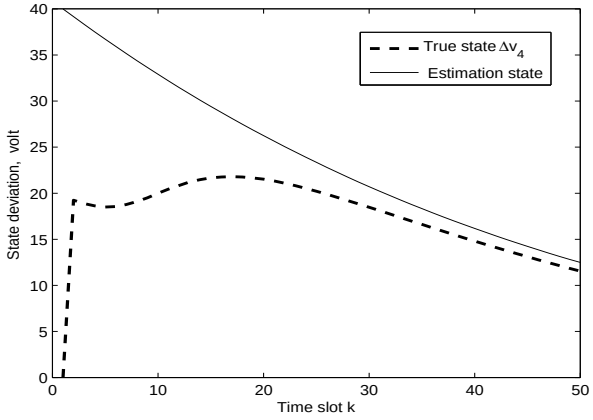


Fig. 18. Δv_3 comparison between the true and estimated state using 8 sensors.

References

[1] X. Zhang, W. Pei, W. Deng, Y. Du, Z. Qi, and Z. Dong, "Emerging smart grid technology for mitigating global warming," *International Journal of Energy Research*, vol. 39, no. 13, pp. 1742–1756, 2015.

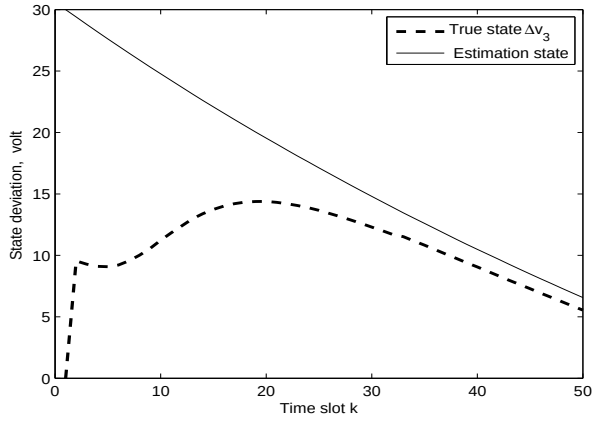


Fig. 19. Δv_4 comparison between the true and estimated state using 8 sensors.

[2] R. Mao and H. Li, "Nobody but you: Sensor selection for voltage regulation in smart grid," *arXiv preprint arXiv:1103.5441*, 2011.

[3] H. Liang and W. Zhuang, "Stochastic modeling and optimization in a microgrid: A survey," *Energies*, vol. 7, no. 4, pp. 2027–2050, 2014.

[4] H. Liang, A. Abdrabou, and W. Zhuang, "Stochastic information management for voltage regulation in smart distribution systems," in *Proc. of the INFOCOM*. IEEE, 2014, pp. 2652–2660.

[5] X. Wang and Q. Liang, "Stabilizing the power supply in microgrid using sensor selection," in *Proc. of the of the Global Communications Conference*, 2012, pp. 3513–3518.

[6] N. Kayastha, D. Niyato, E. Hossain, and Z. Han, "Smart grid sensor data collection, communication, and networking: A tutorial," *Wireless Communications and Mobile Computing*, 2012.

[7] A. P. S. Meliopoulos, G. J. Cokkinides, R. Huang, E. Farantatos, S. Choi, Y. Lee, and X. Yu, "Smart grid technologies for autonomous operation and control," *IEEE Transactions on Smart Grid*, vol. 2, no. 1, pp. 1–10, 2011.

[8] Y. Wang, P. Yemula, and A. Bose, "Decentralized communication and control systems for power system operation," *IEEE Transactions on Smart Grid*, December 2014.

[9] P. Chavali and A. Nehorai, "Distributed power system state estimation using factor graphs," *IEEE Transactions on Signal Processing*, vol. 63, no. 11, pp. 2864–2876, 2015.

[10] Y. Li, "Fully distributed state estimation of smart grids," in *Proc. of the International Conference on Communications*, 2012, pp. 6580–6585.

[11] Y. Weng, R. Negi, and M. D. Ilic, "Graphical model for state estimation in electric power systems," in *Proc. of the International Conference on Smart Grid Communications*, 2013, pp. 103–108.

[12] B. Ruffer, C. M. Kellett, P. M. Dower, and S. R. Weller, "Belief propagation as a dynamical system: the linear case and open problems," *IET Control Theory and Applications*, vol. 4, no. 7, pp. 1188–1200, 2010.

[13] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. R. Kschischang, "The factor graph approach to model-based signal processing," *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1295–1322, 2007.

[14] Y. Hu, A. Kuh, A. Kavcic, and D. Nakafuji, "Real-time state estimation on micro-grids," in *Proc. of the International Joint Conference on Neural Networks*, 2011, pp. 1378–1385.

[15] P. K. Ray and B. Subudhi, "Ensemble Kalman filter based power system harmonic estimation," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 12, pp. 3216–3224, 2012.

[16] M. Yun and B. Yuxin, "Research on the architecture and key technology of internet of things (IoT) applied on smart grid," in *Proc. of the International Conference on Advances in Energy Engineering*, 2010, pp. 69–72.

[17] Q. Chi, H. Yan, C. Zhang, Z. Pang, and L. Da Xu, "A reconfigurable smart sensor interface for industrial WSN in IoT environment," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, pp. 1417–1425, 2014.

[18] I. Bojanova, G. Hurlburt, and J. Voas, "Imagineering an internet of anything," *Computer*, no. 6, pp. 72–77, 2014.

[19] J. Huang, Y. Meng, X. Gong, Y. Liu, and Q. Duan, "A novel deployment scheme for green internet of things," *Internet of Things Journal*, vol. 1, no. 2, pp. 196–205, 2014.

[20] J. Gaoa, Y. Xiao, J. Liu, W. Liang, and P. Chenc, "A survey of communication and networking in smart grids," *Future Generation Computer Systems*, vol. 28, pp. 391–404, 2012.

[21] A. G. Exposito, A. Abur, A. D. L. V. Jaen, and C. G. Quiles, "A multilevel state estimation paradigm for smart grids," *Proc. of the IEEE*, vol. 99, no. 6, pp. 952–976, 2011.

[22] H. Li, L. Lai, and H. V. Poor, "Multicast routing for decentralized control of cyber physical systems with an application in smart grid," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 6, pp. 1097–1107, 2012.

[23] H. Li, F. Li, Y. Xu, D. T. Rizy, and J. D. Kueck, "Adaptive voltage control with distributed energy resources: Algorithm, theoretical analysis, simulation, and field test verification," *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1638–1647, 2010.

[24] E. Ghahremani and I. Kamwa, "Online state estimation of a synchronous generator using unscented Kalman filter from phasor measurements units," *IEEE Transactions on Energy Conversion*, vol. 26, no. 4, pp. 1099–1108, 2011.

[25] D. Buchstaller, E. C. Kerrigan, and G. A. Constantinides, "Sampling and controlling faster than the computational delay," *IET control theory and applications*, vol. 6, no. 8, pp. 1071–1079, 2012.

[26] D. Simon, *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. New Jersey: John Wiley and Sons, 2006.

[27] M. M. Rana and L. Li, "Kalman filter based microgrid state estimation using the internet of things communication network," in *Proc. of the International Conference on Information Technology-New Generations*, 2015, pp. 501–505.

Md Masud Rana is in school of Electrical, Mechanical and Mechatronic Systems at University of Technology Sydney, Australia. His research interests are in the theoretical and algorithmic studies in signal processing and optimizations, statistical learning and inferences for high dimensional data, distributed optimizations and adaptive algorithms, as well as their applications in communications, networked systems and smart grid.

Li Li received Ph.D degree from the University of California, USA. He is servicing as a senior lecturer in school of Electrical, Mechanical and Mechatronic Systems at University of Technology Sydney, Australia. His current research interests are robust control systems, distributed model predictive control of power systems, model reduction of power systems, control on microgrids, vehicle-to-grid coordination method and smart grid market.

A Distributed Cloud based Video Storage System with Privacy Protection

Kang Il Choi, Jung Hee Lee, Bhum Cheol Lee

Network Computing Convergence Lab., ETRI, Daejeon, Korea

forerunner@etri.re.kr, jhlee@etri.re.kr, bclee@etri.re.kr

***Abstract*— In this paper, we present a novel method to protect privacy information for the Cloud based Video Storage System (CVSS), which not only distribute the CVSS geographically, but also use different privacy protection key algorithm for different CVSS to encode/decode the actual video data itself stored in the storage with the subscriber key for the privacy protection so that the system even under the hacking of the CVSS systems, still provide the privacy protection of the video data. We present how this system stores the video stream data transferred from the network connected camera such as IP CCTV. As the system receives the video stream, it masks the privacy related part, such as facial information or car plate number and so on, of the video stream data with a scrambling key and it also encrypts the scrambled video stream data with an encryption key. In this paper, we also present how the system retrieves the privacy information protected video image requested by either the network control centre or end users. When the network control centre or end user requests the networked video, it retrieves the corresponding video image from the video storage first. Then, it decrypts the image with the decryption key and unscrambles the decrypted image with the unscrambling key. Then it transfers the network video back to the network control center or to the end-user. We present the architecture of the distributed CVSS with the privacy protection for the Cloud based Network Function Virtualization System. We also provide flowchart for receive/transmit operation of the DCVSS. Finally, we present a POC implementation of the distributed CVSS in the Cloud based Network Function Virtualization System.**

***Keyword*—** Cloud Video Storage System, Privacy Protection

I. INTRODUCTION

In general, components of a video data storage system, a camera, a monitor, a video recording device, are connected to each other over a closed network to monitor the facility such as a predetermined building or convenience facility.

In this case, a user agent is present in a user device to provide video data to the user device. Accordingly, a camera, a server, and the user device control the camera through mutual interface communication.

Also, video data photographed from the camera is recorded in a storage device within the server and is also transferred to a viewer of the user device.

The above system is generally configured as a closed network among the camera, the server, and the user device for controlling the camera and storing the video data.

In the closed network, the user agent operating in the user device obtains state information of the camera by transmitting a command to the camera through the server, and the camera transfers current state information to the user agent through the server.

A camera server provides interface communication between the user agent and the camera. Also, the camera server enables a user to store an image through the user agent or enables the server to execute a command, such as displaying the image on the viewer in lieu of the user.

When the user desires the image to be displayed on the viewer of the user device, the image transferred to the server is transferred to the user device.

When the user desires to store the image in the storage device, the image transferred to the server is stored in the storage device within the server.

However, a camera monitoring system connected using the closed network has constraints in extensibility. In addition, when the user is not directly connected to the closed network, it is difficult to provide a function capable of controlling the camera through a general network (for example, the Internet) or at least viewing or storing an image.

In this environment, providing the privacy protection is one of the social issues because of several hacking accident ends up with huge privacy information leaking. There are several different approaches to cope with this privacy protection for the video data storage system. However, almost all approaches focus on the IP camera side to provide secure transmit of the video data collected by the IP camera [3-12].

In this paper, we propose a novel privacy protection approach in the video data storage server side, which encoding/decoding the actual video data itself stored in the storage with the subscriber key for the privacy protection so that the system even under the hacking of the CVSS system, still provide the privacy protection of the video data.

The paper is structured as follows. Section II describes an overview of the Cloud based Video Storage System (CVSS) which supports Privacy Protection. We also describe the operation of the CVSS with flowcharts so that how the system

Manuscript received January 29, 2016. This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIP) (B0101-15-233, Smart Networking Core Technology Development)

Kang Il Choi is with the Electronics Telecommunication Research Institute, Daejeon, South Korea (phone: +82-42-860-1704; mobile: +82-10-3861-7758; fax: +82-42-860-5213; e-mail: forerunner@etri.re.kr).

Jung Hee Lee is with the Electronics Telecommunication Research Institute, Daejeon, South Korea (e-mail: jhlee@etri.re.kr).

Bhum Chul Lee is with the Electronics Telecommunication Research Institute, Daejeon, South Korea (e-mail: bclee@etri.re.kr).

supports the privacy protection for the CVSS. In Section III, we describe the architecture of the distributed CVSS (DCVSS) with the Privacy Protection for the Cloud based Network Function Virtualization System. We also describe the operation of the DCVSS with flowcharts so that how the system supports the privacy protection for the DCVSS. Finally, In Section IV, we describe the POC implementation of the DCVSS with the Privacy Protection in the Cloud based Network Function Virtualization System.

II. CLOUD BASED VIDEO STORAGE SYSTEM

Fig. 1 is a control block diagram illustrating a control configuration of a Cloud based Video Storage System (CVSS) [1,2].

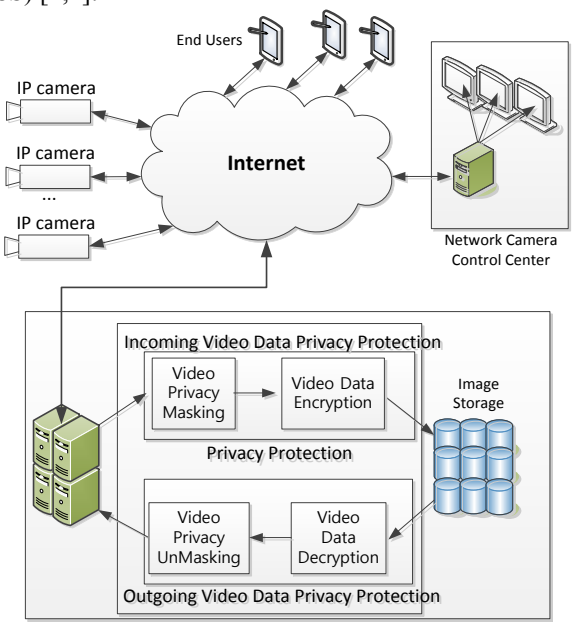


Fig. 1 Cloud Based Video Storage System

In Fig. 1, the CVSS system consists of an "Incoming Video Data Privacy Protection (IVDPP)", an "Outgoing Video Data Privacy Protection (OVDPP)" block and Image Storage. The IVDPP block creates a protected image by applying a privacy protection algorithm to an original image received in the IVDPP block. The IVDPP block stores the protected image into the Image Storage. The OVDPP block retrieves a protected image from the Image Storage and creates the original image by applying a privacy release algorithm to the protected image.

In Fig. 1, the CVSS may be included in a device capable of photographing an image, for example, an imaging device such as a network camera, a mobile communication terminal, and a closed-circuit television (CCTV), or may be stored in a storage server, for example, a network server and a cloud configured to store an image photographed by the imaging device.

In Fig. 1, in the incoming direction, the IVDPP block includes "Video Privacy Masking (VPM)" block and "Video Data Encryption (VDE)" block.

In Fig. 1, the VPM block detects the privacy image from the original image which is received from the remote video capturing system. The VPM block detects the privacy image based on at least one of the edge information of the original image, shape information, color analysis information, and learning information about a previous privacy image.

In Fig. 1, the VPM block obtains the edge information by analyzing the original image and detects a boundary between a face and a portion excluding the face using the edge information.

It determines whether the face is present within the original image by comparing the results of analyzing the original image and shape information about a facial shape. Also, the VPM block compares the color analysis information with unique color distribution information of the face and thus, may more accurately detect the face.

Additionally, the VPM block detects a facial area using learning information about the previous privacy image, including an Adaboost learning scheme.

In Fig. 1, the VPM block performs masking by scrambling the privacy image detected by the VPM block to a privacy protection image using a set of scrambling keys. The VPM block scrambles the privacy protection image to be expressed using a predetermined color and a predetermined figure.

In Fig. 1, the VDE block creates a protected image in which the privacy protection image input from the VPM block and the general image is encrypted using an encryption key.

In Fig. 1, the VDE block may increase security about the privacy protection image by encrypting the privacy protection image and the general image. Then, the VDE block transfers the protected image to the image storage.

The image storage block may assign a unique number capable of recognizing the protected image input from the VDE block and thereby stores the protected image.

The image storage transfers the protected image to the OVDPP block in response to a control command of the Network Camera Control Center or the End Users.

In Fig. 1, the OVDPP block operates relatively inversely to the IVDPP block. In Fig. 1, in the outgoing direction, the OVDPP block includes a Video Data Decryption (VDD) block and a Video Privacy Unmasking (VPU) block.

In Fig. 1, the VDD block decrypts the privacy protection image and the general image from the protected image using a decryption key when the protected image is input from the image storage. Here, the decryption key may be identical to the encryption key, or may be another key capable of performing decryption and corresponding to the encryption key.

In Fig. 1, the VPU block detects the privacy protection image from the privacy protection image and the general image decrypted by the VDD block. The VPU block detects, from the privacy protection image and the general image, the privacy protection image that is expressed using at least one of the predetermined color and the predetermined figure.

The VPU block detects the privacy protection image based on at least one of the edge information of the original image including the privacy protection image and the general image, shape information, color analysis information, and learning information about a previous privacy image. In addition, the VPU block detects the privacy protection image as an image expressed using at least one of the predetermined colors and the predetermined figure of the privacy protection image.

In Fig. 1, the VDD block unscrambles the privacy protection image detected by the VPU block to the privacy image using an unscrambling key. The unscrambling key may be identical to the scrambling key, or may be another key corresponding to the scrambling key.

In response to a control command of the Network Camera Control Center or the End Users, the VDS block transmits the

original image including the privacy image and the general image, or may transfer the original image to the Network Camera Control Center or the End Users.

A. Receive Operation of the CVSS

Fig. 2 is a flowchart illustrating a receive operation method of a Cloud Based Video Storage System (CVSS).

In Fig. 2, the CVSS determines whether a privacy image is detected from an original image when the original image is input, and scrambles the privacy image to a privacy protection image using a scrambling key when the privacy image is detected.

The original image may include at least one of the privacy image including user information, for example, at least one of facial information, license plate information, and privacy information, and a general image excluding the privacy image. Specifically describing, the CVSS determines whether the privacy image is detected from the original image when the original image is input.

In Fig. 2, the CVSS detects the private image based on at least one of the edge information of the original image, shape information, color analysis information, and learning information about a previous privacy image.

In Fig.2, the CVSS obtains the edge information by analyzing the original image M1, may detect a boundary between a face and a portion excluding the face using the edge information, and may determine whether the face is present within the original image by comparing a result of analyzing the original image and shape information about a facial shape.

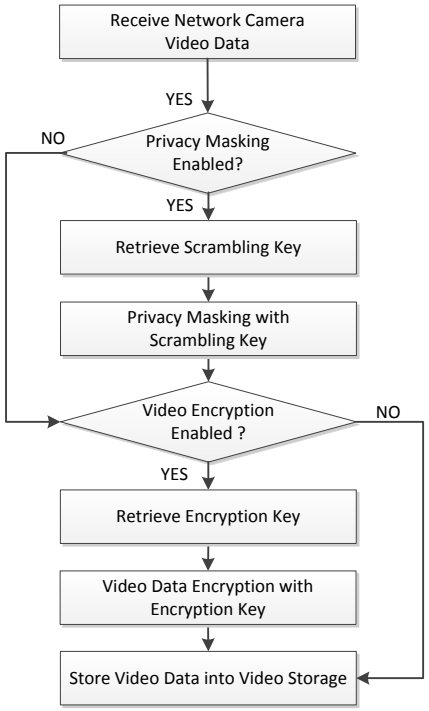


Fig. 2 Receive Operation of the Cloud based Video Storage System

Also, the CVSS compares the color analysis information with unique color distribution information of the face and thus, may more accurately detect the face. Additionally, the CVSS may detect a facial area using learning information about the previous privacy image, including an Adaboost learning scheme.

When the privacy image is detected, the CVSS performs masking by scrambling the detected privacy image to the privacy protection image using a set of scrambling key.

After this operation, the CVSS creates a protected image in which the privacy protection image and the general image are encrypted using an encryption key, and transfers the protected image to the image storage.

When the privacy image is not detected from the original image, the CVSS performs encryption operation on the original image.

B. Transmit Operation of the CVSS

Fig. 3 is a flowchart illustrating an operation method of a CVSS when an image is output from the CVSS.

In Fig. 3, when an image request signal requesting an original image is input, the CVSS receives a protected image stored in the image storage, and the CVSS decrypts the protected image to an original image including a privacy protection image and a general image using a decryption key and detects the privacy protection image.

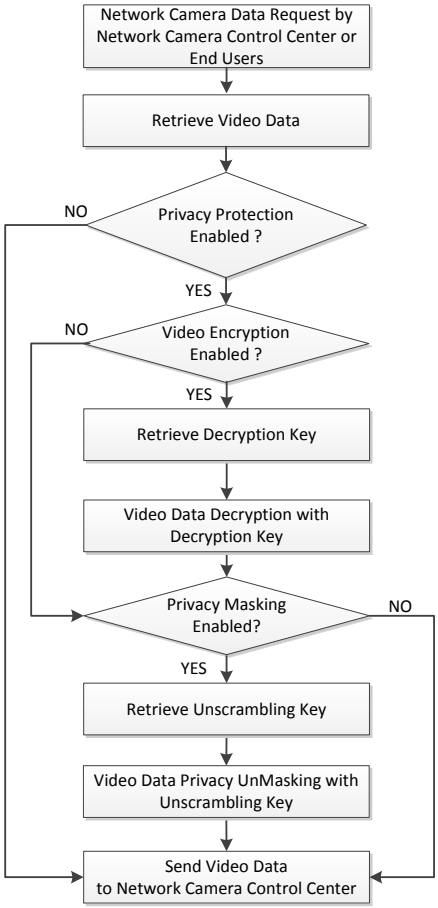


Fig. 3 Transmit Operation of the Cloud based Video Storage System

In response to a control command of the Network Camera Control Center or the End Users and the image request signal requesting the original image, the CVSS receives, from the image storage, the protected image to which a privacy protection algorithm corresponding to the original image is applied among images stored in the image storage.

Next, the CVSS decrypts the protected image to the original image which is including the privacy protection image and the general image using a set decryption key.

The CVSS detects, from the original image, the privacy protection image that is expressed using at least one of a predetermined color and a predetermined figure.

After this operation, the CVSS unscrambles the privacy protection image to the privacy image using an unscrambling key, and creates the original image including the privacy protection image and the general image.

The CVSS creates the privacy image by unscrambling, using the unscrambling key, the privacy protection image scrambled using a scrambling key.

Next, the CVSS creates the original image including the privacy protection image and the general image, and transfers the original image to the Network Camera Control Center or the End Users or a predetermined device having input the image request signal.

III. Distributed Cloud based Video Storage System

Fig. 4 is an architecture diagram illustrating a distributed Cloud based Video Storage System (DCVSS).

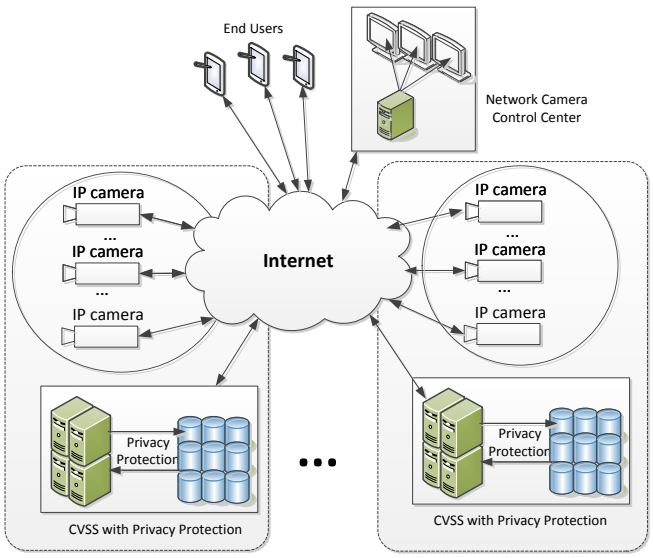


Fig. 4 An architecture of the Distributed CVSS

In Fig. 4, each individual DCVSS process the video images transmitted by a group of IP cameras, which is geographically close to the DCVSS, so that it can provide privacy protection for the video images generated by the group of the IP cameras.

In Fig.4, the Network Camera Control Center configure which DCVSS will handle which IP cameras in terms of the CVSS configuration and IP camera's configuration, which means the Network Camera Control Center knows who process the specific IP cameras video image and provide the privacy protection for the IP cameras.

When an end user request their own image, the network camera control center re-route the end-users request to the DCVSS who actually processed the image so that the DCVSS, who actually process the end user's IP camera, respond to the end-user's request.

In Fig. 4, by distributing the DCVSS system geographically so that the DCVSS process the video images transmitted by a group of IP cameras which is geographically close to the DCVSS, we can provide the Quality of Service requirement from the end-user, such as low latency for the video image processing.

In Fig. 4, by applying separate privacy protection key algorithm (which arithmetically identical but generate different key output) per geographically separated DCVSS, we can also provide separation of privacy protection in case of hacking, so that hacking on the one DCVSS doesn't affect the other DCVSS.

A. *Receive Operation of the DCVSS*

Fig. 5 is a flowchart illustrating a receive operation method of a DCVSS. In Fig. 5, if privacy protection and privacy masking is enabled, then the DCVSS retrieve both personal scrambling key and site dependent scrambling key. And the DCVSS do privacy masking the video data with combined (personal + site) scrambling key. If video encryption is enabled, the DCVSS retrieve both personal encryption key and site dependent encryption key. And the DCVSS encrypt the video data with combined (personal + site) encryption key.

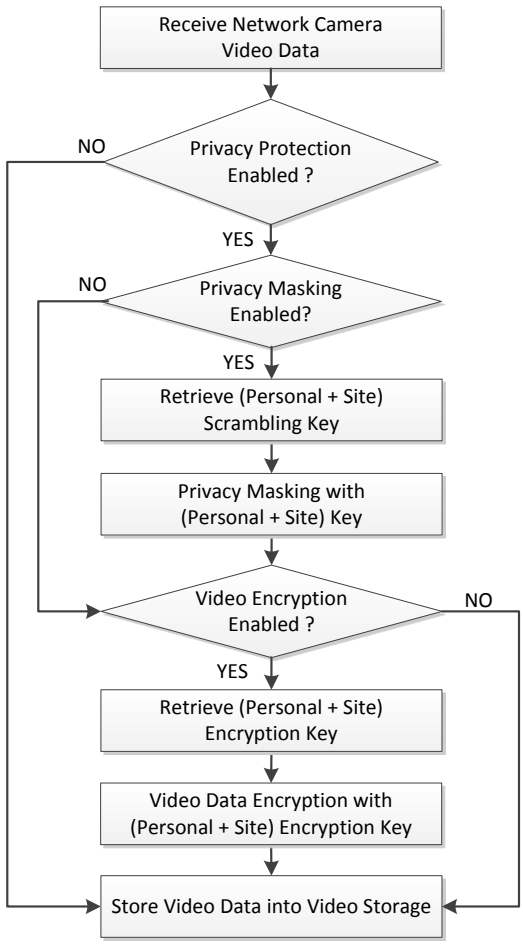


Fig. 5 Receive Operation of the DCVSS

B. *Transmit Operation of the DCVSS*

Fig. 6 is a flowchart illustrating transmit operation method of a DCVSS. In Fig. 6, if privacy protection and video encryption is enabled, then the DCVSS retrieve both personal decryption key and site dependent decryption key. And the DCVSS decrypt the video data with combined (personal + site) decryption key. If privacy masking is enabled, the DCVSS retrieve both personal unscrambling key and site dependent unscrambling key. And the DCVSS unmasking the video data with combined (personal + site) unscrambling key.

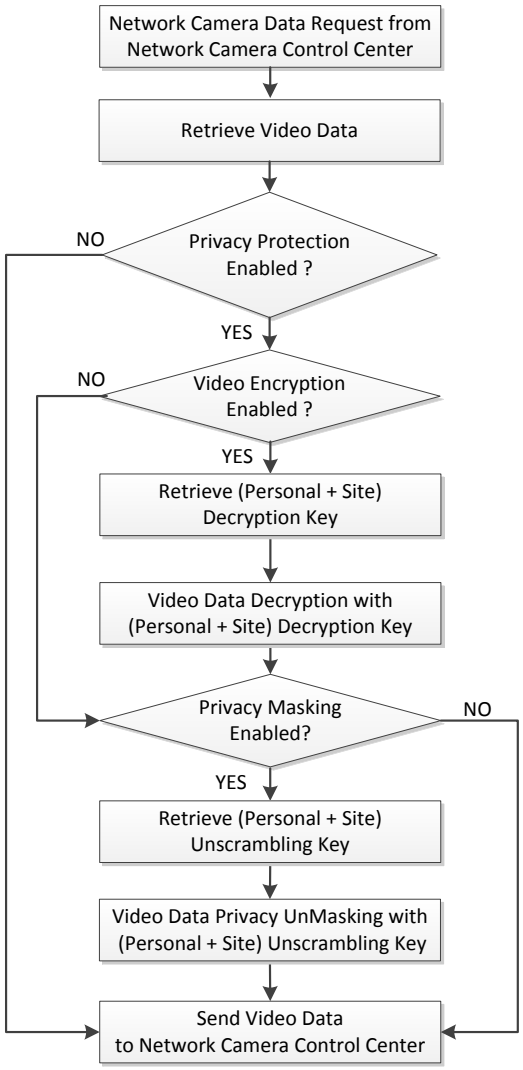


Fig. 6 Transmit Operation of the Cloud based Video Storage System

IV. An POC Implementation of the Distributed CVSS in the Cloud Based Network Function Virtualization System

Fig. 7 is a POC implementation of the distributed Cloud based Video Storage System (DCVSS) in the Cloud based Network Function Virtualization System, which consist of a DCVSS Control Center, Daejeon DCVSS and Seoul CVSS.

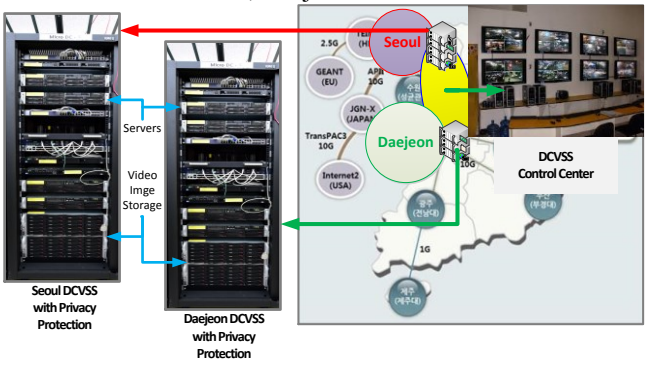


Fig. 7 An implementation of the Distributed CVSS in the Cloud Network Function Virtualization System

In Fig. 7, each DCVSS consists of Servers, which process the IP camera network image, Video Image Storage, which stores the IP camera network images, and Privacy Protection Module, which running on the Servers, encrypts and decrypts the IP camera images for privacy protection.

In Fig. 7, each DCVSS also apply separate key system for the privacy protection algorithms. End users from Daejeon areas access their video image through Daejeon DCVSS with privacy protection, and the End users from Seoul area access their video image through the Seoul CVSS.

In Fig. 7, there are two DCVSS system with privacy protection, Daejeon DCVSS, which provide privacy protection for the video images generated by the group of the IP cameras located in the Daejeon area, and Seoul DCVSS, which provide privacy protection for the Seoul areas

In Fig. 7, the DCVSS Control Center control the setup process for each DCVSS and configure each DCVSS so that each DCVSS not only process the video images from the group of IP cameras located geographically close to the DCVSS but also use different privacy protection key algorithm. The Daejeon DCVSS configured to process the group of IP cameras from the Daejeon area with privacy protection key algorithm A, and the Seoul DCVSS configured to process the group of IP cameras from the Seoul area with privacy protection key algorithm B. With this configuration, even the Daejeon DCVSS hacked so that the video image of it hacked, the video images in the Seoul DCVSS are safe and provide privacy protection because these two DCVSS use different privacy protection key algorithm.

V. Conclusion

In this paper, we presented a novel method to protect privacy information for the Cloud based Video Storage System (CVSS), which encoding/decoding the actual video data itself stored in the storage with the subscriber key for the privacy protection so that the system even under the hacking of the CVSS system, still provide the privacy protection of the video data. On the top of this, by distributing the CVSS system geographically and by applying separate privacy protection key algorithm, we can prevent huge privacy information leaking and localize the privacy leaking. We presented how this system stores/retrieves the video stream data transferred from the network connected camera such as IP CCTV to/from the cloud based video storage system. In the paper, we presented the detailed procedures and algorithms for these processes. In this paper, we masked the privacy related part and encrypted the scrambled video with an encryption key so that we protected the privacy information of the networked video stream, which is stored in the cloud based video storage system. We presented the architecture of the distributed CVSS with the privacy protection for the Cloud based Network Function Virtualization System. We also provided flowchart for receive/transmit operation of the DCVSS. Finally, we presented the implementation of the distributed CVSS in the Cloud based Network Function Virtualization System

References

[1] Kang Il Choi, Jung Hee Lee and Bhum Cheol Lee, "Cloud based Video Storage System with Privacy Protection", *ICACT2015, 2015, pp. 448–451.*

[2] Kang Il Choi, Bhum Cheol Lee, Seung Woo Lee, Young Ho Park, Jung Hee Lee and Sang Min Lee, "Image processing apparatus and operation method thereof", *US20150055775 A1, Aug 20,2013*

[3] DJ Neal and Syed Shawon Rahman, "VIDEO SURVEILLANCE IN THE CLOUD?", *International Journal on Cryptography and Information Security (IJCIS),Vol.2, No.3, September 2012*

[4] Yong-Hua Xiong, Shao-Yun Wan, Yong He and Dan Su, "Design and Implementation of a Prototype Cloud Video Surveillance System", *Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol.18, No.1 pp. 40-47, 2014*

[5] Biao Song, Yuan Tian and Bingyin Zhou, "Design and evaluation of remote video surveillance system on private cloud", *2014 4th International Symposium on Biometrics and Security Technologies, 26 August 2014*

[6] Chia-Feng Lin, Shyan-Ming Yuan, Muh-Chyi Leu and Ching-Tsorng Tsai, "A framework for scalable cloud video recorder system in surveillance environment", *IEEE 9th International Conference on Ubiquitous Intelligence and Computing and IEEE 9th International Conference on Autonomic and Trusted Computing, UIC-ATC 2012, Article number 6332062, Pages 655-660*

[7] Räty, T.D., "Survey on contemporary remote surveillance systems for public safety", *IEEE Transactions on Systems, Man and Cybernetics, art. no. 5422679, pp. 493-515*

[8] Kim, I.S., Choi, H.S., Yi, K.M., Choi, J.Y., Kong, S.G., "Intelligent visual surveillance - A survey", *International Journal of Control, Automation and Systems, 8 (5), pp. 926-939*

[9] Zhao, Z., Cui, X., Zhang, H., "Cloud storage technology in video surveillance", *Advanced Materials Research 2012, 532-533, pp. 1334-1338*

[10] Rodríguez-Silva, D.A., Adkinson-Orellana, L., González-Castaño, F.J., Armiño-Franco, I., González-Martínez, D., "Video surveillance based on cloud storage", *2012 IEEE 5th International Conference on Cloud Computing, CLOUD 2012, art. no. 6253615, pp. 991-992.*

[11] Hossain, M.S., Hassan, M.M., Qurishi, M.A., Alghamdi, A., "Resource allocation for service composition in cloud-based video surveillance platform", *Proceedings of the 2012 IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2012, art. no. 6266418, pp. 408-412*

[12] Pearson, S., "Taking account of privacy when designing cloud computing services", *2011 IEEE 3rd International Conference on Communication Software and Networks, ICCSN 2011, art. no. 6014715, pp. 245-249*

Kang Il Choi received B.S. degree in Computer Science from KAIST, Korea and M.S. degree in Computer Science from Sogang University in 1992 and 1994, respectively. He is currently senior researcher of Electronics and Telecommunications Research Institute (ETRI), Korea. His research interests are Multicore Parallel Processing, Distributed Cloud Data Center, Data Plane Acceleration Technology and Network Virtualization.

Jung Hee Lee received B.E. and M.S. in Electronic Engineering at Kyungpok National University in 1984 and 1991, respectively. She is currently principal researcher of Electronics and Telecommunications Research Institute (ETRI), Korea. Her research interests are Flow based Network Processor, Multicore Parallel Processing, High Speed Parallel Switching and Network Virtualization.

Bhum Cheol Lee received M.S. and Ph.D. degree in Electric Engineering from Yonsei University, Korea in 1983 and 1997, respectively. He is currently Manager of Networking Computing Convergence Lab. in Electronics and Telecommunications Research Institute (ETRI), Korea. His research interests are Smart Network, Parallel Flow Processing and Network Virtualization.

A Flexible FPGA-to-FPGA Communication System

An Wu*, Xi Jin*, Xueliang Du**, ShuaiZhi Guo*

*Department of Physics, University of Science and Technology of China, Hefei, Anhui Province, China

**Department of System verification, Chinese Academy of Science Institute of Automation, Beijing, Beijing Province,China

wuan@mail.ustc.edu.cn, jinxi@ustc.edu.cn, xueliang.du@ia.ac.cn

***Abstract*—In high-performance computing systems, each computing node communicates via a high-speed serial bus to ensure sufficient data transfer bandwidth. However, each computing node of different bus protocols is very difficult to communicate directly, which is not conducive to the extensibility of HPC (High performance computing) clusters. In this paper, we propose UPI, a inter-node communication interface based on FPGA, which can transmit different bus protocols (PCIe protocol and Ethernet protocol) simultaneously. More importantly, many different bus-supported computing nodes can be connected to the same HPC system. We implemented our UPI system on "Gemini" prototype verification board with two Xilinx Virtex-6 FPGAs. The results show that the transmission speed of the UPI can reach 11.04Gpbs (PCIe Gen2 X4) and 4.32Gpbs (Gigabit Ethernet) when DMA payload sizes is greater than 260KB and 80KB, respectively.**

***Keyword*—FPGA-based SoCs, PCIE, Gigabit Ethernet, HPC**

I. INTRODUCTION

With the rapid development of high-performance FPGA-based devices, high-performance computing system performance bottleneck has shifted from the ability of single node to the architecture of the HPC clusters. In the system-level or board-level interconnect system, high-speed serial bus technology with its enormous advantage is rapidly replacing traditional parallel bus technology and becoming the main technology of high-speed FPGA-based SoC design. As the applications of the high-speed serial bus technology gradually expand to all research area, more and more articles focus on the technology, especially the multi-node interconnect technology and the high-speed I/O interface technology.

Effective inter-node communication is receiving significant attention due to its increasingly important applications in high-performance computing, thus inter-node interconnection technology is drawing lots of attractions from more and more researchers. The interconnection methods are based on three main protocols: PCIe, Ethernet, and Serial RapidIO (SRIO). However, they have their own using domain. They can't communicate each other directly, and for the same reason, each computing node can't be compatible on the means of communication.

In order to achieve a good compatibility between these protocols, many problems cannot be neglected. For example, different bus communication requires bridging conversion, the process is complicated and performance loss is hard to avoid. The problem appears in these bus controllers' hardware design. In FPGA-based SoCs, both of Ethernet controller and PCIe controller have an external physical layer (PHY) chip. The controllers' PHY chip of mentioned three types are different, mainly due to its PHY's interface signals and communication speeds are not identical. Ethernet PHY chip mainly realizes 64b/66b encoding (Gigabit Ethernet), and PCIe PHY chip completes the 8b/10b encoding (PCIe 2.0 and below). However, they still have a lot in common, e.g., scrambling and parallel-to-serial conversion all need to be implemented in PHY chip.

The simplest way to solve the compatibility issue is to add all these bus protocols to a computing node, but it requires more PHY chips, differential pairs and hardware resources. A better approach is to communicate with these computing nodes through the same interface without changing the interface types. Generally, the physical layer of PCIe, Ethernet and SRIO cannot be shared. We try to merge the physical layer of three bus protocols, finally, we built a unified physical layer interface which is actually the same functions for their upper layer protocols.

The benefits of a unified physical layer design are as follows:

- A unified physical layer can provide a compatible interconnection bus with three protocols.
- Either PCIe, Ethernet or SRIO devices can be plugged into the same interface for communication.
- The problem of performance loss in bridging process can be settled.

In this paper, we proposes UPI (Unified PHY Interface) system, a flexible interconnection system of inter-node communications based on FPGA devices. Our design shares

Manuscript received February 24, 2016. This work is a follow up of the invited journal of the accepted conference paper for the 18th International Conference on Advanced Communication Technology. This research was supported by the "Strategic Priority Research Program" of the Chinese Academy of Sciences, Grant No.XDA06010402-4.

An Wu is with the Department of Physics, University of Science and Technology of China, Hefei, Anhui Province, China (corresponding author to provide phone: +86-159-5510-2092; e-mail: wuan@ mail.ustc.edu.cn).

Xi Jin is with the Department of Physics, University of Science and Technology of China, China (e-mil:jinxi@ustc.edu.cn).

XueLiang Du is with the Department of System verification, Chinese Academy of Science Institute of Automation, China (e-mil: xueliang.du@ia.ac.cn).

Shuaizhi Guo is with the Department of Physics, University of Science and Technology of China, China (e-mil:dybjxmg@mail.ustc.edu.cn).

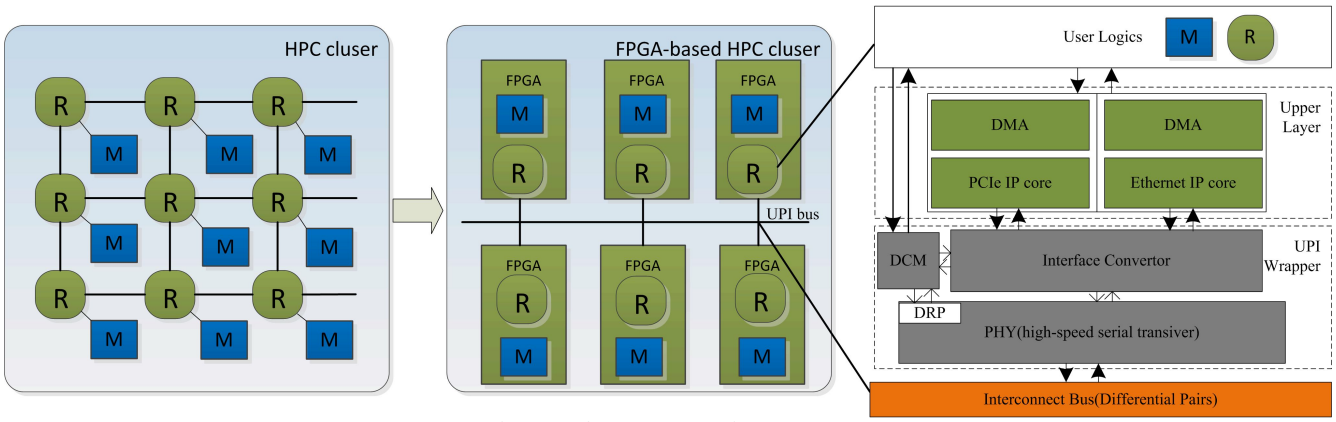


Fig.1. Hardware structure of UPI system

the same portion of PCIe 2.0 and Gigabit Ethernet's physical layers, and merges their different parts. The UPI can transmit PCIe and Ethernet packets with one physical layer chip. The computing node of HPC (High performance computing) clusters using different protocols can connect with each other, meanwhile, the bridging delay and loss in performance can be eliminated, and through the interface design we can implement a more flexible and efficient FPGA-based computing clusters.

The remainder of this paper is organized as follows: background and related work are discussed in section II. We describe the design of UPI in section III. The experiments and results are discussed in Section IV. Finally, we present a conclusion in Section V.

II. Related Work

A. Background on PCIe and Gigabit Ethernet

PCIe is a high-speed serial bus includes transaction layer, data link layer and physical layer. The Transaction layer contains TLP (Transaction Layer Packets) control mechanism. The Data Link layer primary responsibility is to provide a reliable mechanism for exchanging TLPs between the two components on a link [1]. At the physical layer (PHY), the PCIe bus provides a serial high throughput interconnect medium between two devices. PCIe PHY contains two sub-layer: Physical Coding Sub-layer (PCS) and Physical Media Attachment (PMA)[2]. There have been three versions of the PCIe bus. For a single lane, data transfer rate for versions 1.x, 2.x and 3.x are 2,4 and 8Gbps [3].

There are two layers in hardware of Gigabit Ethernet, including physical layer and data link layer (DLL). The main function of DLL is to complete the frame transmission and frame reception. Gigabit Ethernet PHY has three main functions: First, It's provide transferring path of data to data terminal equipment; Second, to be a proper entities for data transmission, not only to ensure that data transfers properly on it, but also to provide sufficient bandwidth and reduce channel congestion; third, complete management of PHY. Upper layer and PHY interconnect with each other via a MII/GMII/RGMII/SGMII interface, through the manage interface in MII, the upper layer can control and monitor PHY [4].

B. Xilinx GTX

Xilinx GTX is a programmable high-speed serial transceiver capable of speeds from 500Mbps to 12.5Gbps.

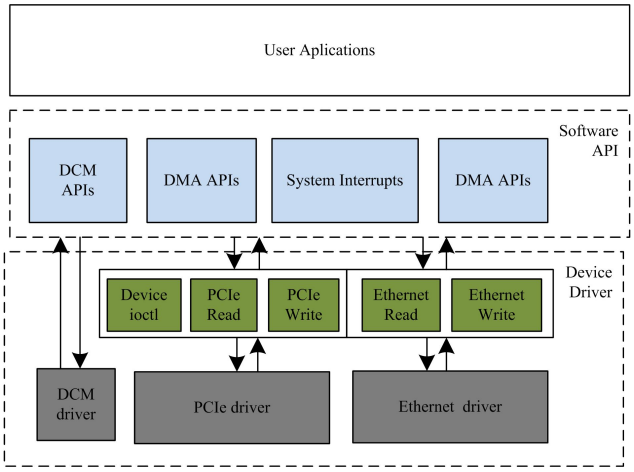


Fig.2. Software structure of UPI system

GTX module in Xilinx FPGA can be realized different serial interconnect protocols, such as SATA, PCIe, EMAC and SRIO. Dynamic Reconfiguration Port (DRP) is an interface module which allows the dynamic change of parameters of the GTX. Through the DRP interface, we can realize the dynamic changes of each interconnect protocol, making it possible for system to adapt to the protocol change. QPLL (Quad PLL) and CPLL (Channel PLL) are two kinds of PLL circuit with different clock rates embedded in GTX module [5].

C. Existing Work

Ethernet is a widely used protocol in HPC computing sysems, which key is its inter-node routing policy. Most system designers make use of the Ethernet protocol to constitute multi-node communication network, the transmission rules follows the Ethernet protocol. However, as more and more high-performance embedded devices appear, researchers are turning their attentions to the direct communication of GPU, FPGA, DSP and other processing units. As a high-speed communication interface, the performance and bandwidth of PCIe meets our design requirements.

Many technologies based on PCIe protocol are proposed, such as InfiniBand and Hypertransport. The InfiniBand Architecture (IBA) is an industry-standard fabric designed to provide high bandwidth/low-latency computing, scalability to ten-thousand nodes and multiple CPU cores per server platform, and efficient utilization of compute processing resources. InfiniBand adapters and switches deliver 56Gb/s bandwidth today and are expected to deliver 100Gb/s by

2016 [6].

For chip-to-chip communications, AMD HyperTransport (HT) are used to connect between CPUs, as well as between CPU and memory. It provides the integral interconnect backbone structure that links all of the core functional units (processor, memory and I/O elements) in a board-level system. As an optimized board-level architecture, HyperTransport provides the lowest possible latency, harmonizes interfaces and supports scalable performance [7].

BlueLink is custom interconnect toolkit for commodity FPGA clusters. Traditional standard protocols such as Ethernet and Interlaken are a boon for FPGA-to-other-system interconnect, they are inefficient and unnecessary for FPGA-to-FPGA interconnect. BlueLink can use all the transceivers available on an FPGA board, over any physical medium. Comparing to 10G Ethernet, 10G BlueLink uses 65% of the logic and registers of 10G Ethernet and uses 15% of the memory of 10G Ethernet. To consider throughput, BlueLink's latency is about equivalent to Ethernet in the fully-loaded case.

III. System Design

In this section we illustrate UPI hardware architecture and software API, as well as explaining the flexibility and compatibility of UPI architecture.

A. Overall System Architecture

Classic HPC cluster structure is shown in Figure 1. The letter M indicates the the computing machine, and letter R indicates the routing policy. We implemented this structure on FPGA-based devices as shown in the middle of Figure 1. In our UPI system, the computing machine and routing policy are integrated into FPGA's user logics. Each FPGA-based computing node is communicate with each other through our high-speed UPI bus. The UPI interface is responsible for connecting each FPGA-based node between user logics and UPI bus.

The UPI interface consists of a hardware component and a software component as shown in the right of Figure 1 and Figure 2. The hardware component mainly consists of two different types of bus controllers, the unified FIFO interface, Interface Convertor and the DRP control module. Interface Convertor is used to convert two different physical layer interface signals into a standard GTX interface signals. The DRP control module is used to dynamically reconfigure the parameters of GTX.

Software component is divided into two parts, including standalone board driver and testing code. Our standalone board driver consists of PCIe and Ethernet device driver, DRP control module driver and DMA driver. Test code consists of PCIe TLP packets reading and writing tests, Ethernet TCP/UDP packets reading and writing tests, equipment switching test. To make our architecture more flexible, the software API provides the simplest and effective way to call the underlying driver functions, as well as shielding the details of low-level operations. By switching the functions of high-speed serial transceiver, user logics can easily achieve the mutual communication between the two protocols.

B. DRP Control Module

DRP control module (DCM) needs to complete link speed selection task and link training control task. Thus, two important parameters for DCM to consider are link width and link data rate. After system boot up, PCIe requires link training process to negotiate the link width and link speed between two sides of PCIe controller. Ethernet also has a similar link training process. The key function of DCM is a link training state machine, which stores the link state of two protocols. When the link is switched on, DCM stores current link state and jump to the next state without need to retrain link. The switching time of the two controller is CPLL reset to CPLL locked. The DCM state machine is shown in Figure 3.

At the beginning of link initialization process, both sides are in Silent mode. Host side starts to seek Device side by sending training sequence in Seek mode. If device side sends its training sequence back, the state will jump to the Discovery mode. In this state, each side sends current link width and link rate for handshaking.

UPI has three configuration modes for link width:

1) *If both sides are four lanes, DCM will jump to ISM_4X_MODE state;*

2) *Theoretically, two lanes have six kinds of interconnect methods. Considering that each channel sends the same packets data, DCM will jump to ISM_2X_MODE state;*

3) *In this mode, DCM will jump to ISM_1X_MODE_LANE3, ISM_1X_MODE_LANE2, ISM_1-X_MODE_LANE1 or ISM_1X_MODE_LANE0 state, the details is as follows.*

The configuration modes can be dynamically changed by ISM_2X_RECOVERY, ISM_1X_RECOVERY or Discovery state. We build a 32-bit width data interface to transmit packet and connect it to Interface Convertor module. Each channel has different negotiation methods corresponding to UPI channel features in mode three.

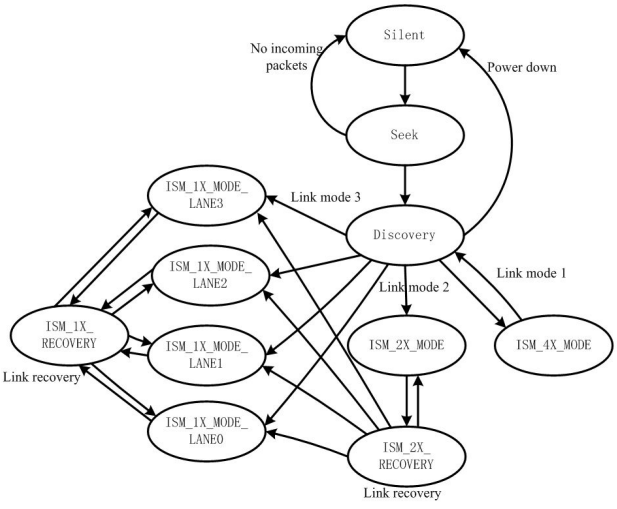


Fig.3. DRP control module link training state machine

The link speed can be changed by modifying QPLL or CPLL multiplication factor and crossover factor. GTX supports 16, 32, 64 and other data width, higher data width can be achieved by stitching several GTXs. In order to complete the link speed and width adjustment, DRP control

module generates different interface signals in different states. For example, when the bus controller is PCIe Gen2 X4, the CPLL reference clock can be adjusted to 125MHz, CPLL output rate 2.5GHz and data link width can be adjusted to 32 bit data with 4bit k symbol indicator.

C. Interface Convertor

PCIe PHY interface is called PIPE interface, and Ethernet PHY interface is called RGMII In order to achieve a mutually compatible interfaces, we need to make PCIe and Ethernet physical layer interface convert to the same GTX interface. This conversion process is completed in Interface Converter. Table 1 shows the interface signals associated with the GTX, including indicators sending and receiving 32-bit words with 4-bit k symbol and link status control signals.

TABLE I
INTERFACE CONVERTOR RELATED SIGNAL

	Signal	I/O	Description
GTX	mac_phy_rxdata[31:0]	I	receive data
	mac_phy_rxdatak[3:0]	I	K character indication
	mac_phy_txdata[31:0]	O	transmit data
	mac_phy_txdatak[3:0]	O	K character indication
	phy_mac_rxvaild[3:0]	O	receive data is valid
	phy_mac_rxelecidle[3:0]	O	receive electrical idle
	phy_mac_phystate[3:0]	O	PHY functions
	mac_phy_txdetectrx_loopback[3:0]	I	enable receiver detec-tion sequence
	mac_phy_txelecidle[3:0]	I	transmit electrical idle
	mac_phy_txcompliance[3:0]	I	Compliance sequence
	macphy txpolarity[3:0]	I	Invert the received data when asserted
	mac_phy_powerdown[2:0]	I	PHY power down
DRP	bus_state_write[5:0]	O	current controller state write to DRP
	bus_state_read[5:0]	I	current controller state read to DRP
	bus_switch_en	O	bus switch enable
	bus_link_speed[3:0]	O	bus link speed
	bus_link_width[15:0]	I	bus link width

The Interface Convertor mainly includes the following features:

- A set of dual-port RAM, the amount of which is equal to the converted upper interfaces.
- Combining and scattering data to meet UPI interface data width.
- Generating PHY's control signals. If the control signals provided by GTX interface, it can be directly connected to GTX. Signals that GTX does not provide will be generated in Interface Converter according to the control signals' relations or be set to a constant value. For example, the PCIe PIPE interface signal mac_phy_blockaligncontrol is only used in PCIe 3.0, which isn't used in our design, so we give it a constant value in Interface Converter.
- Transmitting link information to the DCM module, completing real-time parameter changes for GTX. The link status signals will change into the DCM signals.

Interface Convertor signals are illustrated in Table 1. Some signals, such as clock and reset, are omitted in this table.

D. Data structures

There are three kinds of data structure in our UPI system: Primary data with DMA descriptors, PCIe and Ethernet packets data with their own protocols, GTX data with Control symbol. The Primary data has been stored into DDR with some DMA descriptors. When PCIe or Ethernet controller obtains a writing command from CPU, a segment of primary data will be sent to the controller. PCIe controller transfers primary data to Transaction Layer Packets(TLP). Data Link Layer Packets(DLLP) are in charge of link maintenance. TLP packets will be assembled with a sequence number and a LCRC code in Data Link Layer. Each TLP can accommodate 4096 Bytes data payload. In Ethernet controller, related packets are TCP, UDP, etc., which have 256 Byte's data payload.

When GTX receives packets data form controllers, each kind of packets is marked with a identifier and two kinds of control symbols (CON symbol and END symbol). We distinguish PCIe and Ethernet packets with binary code "01" and "10", respectively. Different packets have different CON symbol. For example, STP symbol is added to the head of TLP packets corresponding to the identifier. The details of CON symbols are listed in table 2. In our design, the packets with identifier is only generated and digested in UPI layer.

TABLE II
CONTROL SYMBOLS IN DCM

Standard	identifier	Control symbol
TLP	01	K27.7
DLLP	01	K28.2
TCP	10	K28.0
UDP	10	K28.4

E. Hardware Interface and Software API

Our UPI system interface includes the following functions:

- FIFO-based DMA interface.
- DCM interface.
- System interrupt interface.

We use FIFO as the DMA interface for three reasons. First, FIFO can be used between two clock domains. Our PCIe system clock is 125MHz while Ethernet is 100MHz. Two clock domains can be isolated by FIFO interface. Second, FIFO is a standard interface. It can shield the low-level details of UPI hardware, so the transmission process in UPI is transparent for user logic. Third, the width and depth of FIFO can be configured by circumstances of the design and it's convenient for developers to comprehend packet structures. FIFO-based DMA interface is used to transmit data from system's memory to system's I/O controller.

The DCM interface uses a simple DCM bus to read and write DCM registers. We provide DCM interface to make our system easier to operate. For some customized HPC clusters, the DCM interface provides a function to change the bus protocol to meet developers' requirements. Through the DCM Interface, developers even can change the type of communication protocol, which makes the system more flexible of protocol switching.

System interrupt interface is a standard vectored interrupt interface. The main functions provided by the system interrupt interface are software interrupt generating, interrupt

source searching, interrupt number generating and interrupt priority setting. Different interrupt numbers mean that different interrupt exceptions and different CPU response process.

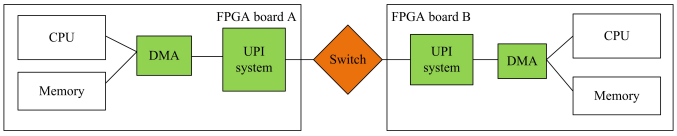


Fig.4. Structure of FPGA-based verification system

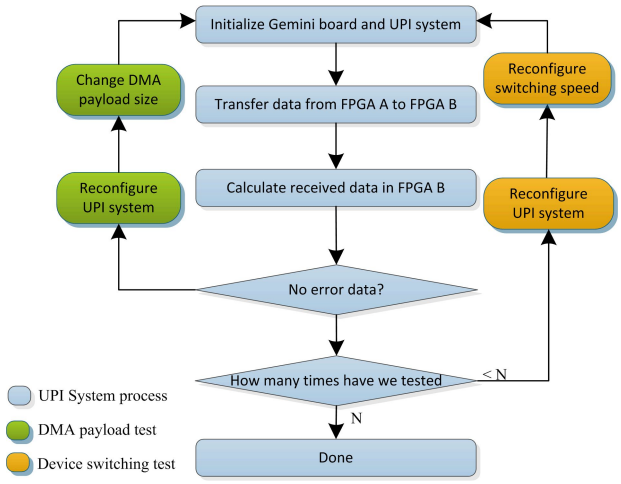


Fig.5. Flow chart of UPI testing program

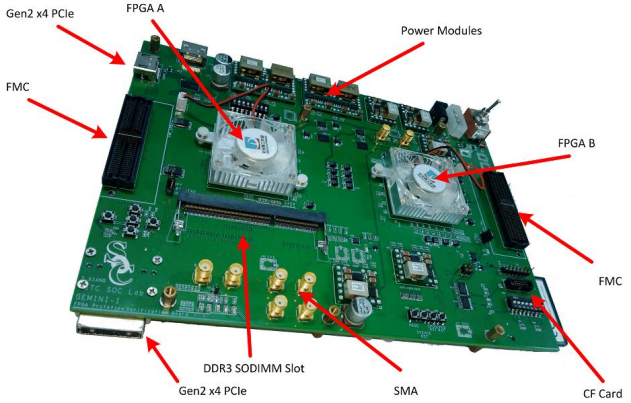
Therefore, a interrupt-supported CPU is required to handle PCIe and Ethernet interrupts. We set these two kinds of interrupts to the same priority, and provide the appropriate interrupt number and interrupt type functions, make CPU to poll interrupt easily after the interrupt exceptions is generated.

Corresponding to the hardware modules, software API is also provides three functions: 1: DMA data read and write. 2: DCM read and write. 3: software interrupts. These APIs are listed as follows, some basic functions such as system reset are omitted in this list:

```
unsigned int drp_write(unsigned char addr, unsigned int data);
unsigned int drp_read(unsigned char addr);

int PCIe_dma(int len, int *ddr_data_1, int *ddr_data_2);
int GMAC_dma(int len, int *ddr_data_1, int *ddr_data_2);

void PCIe_Interrupt(unsigned int ictl_prio,int ictl_number);
void GMAC_Interrupt(unsigned int ictl_prio,int ictl_number);
```



IV. EVALUATION

In this section we implemented UPI system on "Gemini" prototype verification board. Then we evaluate UPI's performance and compare it with other I/O technologies. Finally, we show UPI's flexibility by presenting three practical applications that employ UPI as their communication interface.

A. System Verification Platform

Our system verification platform called "Gemini" is shown in Figure 6. The main hardware components provided by "Gemini" board are two PCIe slots, two Xilinx xc6vlx365tff1156-1 FPGAs, a SODIMM (Small Outline Dual In-line Memory Module) DDR3 and a CF Card's slot. The UPI with a DMA-oriented Synopsys PCIe controller and a Synopsys Ethernet controller is integrated into each FPGA. we also implemented a SoC (system on chip) with a Microblaze CPU. The structure of FPGA-based verification system is shown in Figure 4. The Microblaze CPU communicate with UPI by ARM AXI (Advanced Extensible Interface) bus.

Figure 5 shows the whole process of UPI verification test. Primary data has been stored into DDR with some DMA descriptors. When UPI obtaining a writing command from CPU, a data segment will be sent to UPI system. UPI systemtransfers primary data to physical layer. A typical program for UPI test is as follows:

1) Download bit and elf file to the board.

2) Initialize the CF card, DDR controllers, UPI system.

3) Configure DMA descriptors, move data from CF card into DDR, configure GTX to PCIe gen2.0 X4 mode.

4) Transfer data from FPGA A to FPGA B.

5) Store data in another DDR address in FPGA B.

6) Calculate complete time.

7) Compare the two data segments, calculate error amount if they are different.

8) Switch GTX to Ethernet mode. Repeat steps 3-5.

9) Reconfigure switching speed, and repeat steps 3-6.

B. System test

The UPI simulation process uses standard functional test to simulate UPI system as shwn in Figure 7. The simulation model used in our test is described as follows: 1: UPI_top is the top-level module of UPI system. The gmac, pcie and gpio are some sub-modules in UPI_top module. 2: Our UPI test system have two differential input clock. The host provide 100MHz reference clock and transmit the clock to Gemini board. We use an extra BUF module to receive reference clock before clock is divided in PLL module. After clock is divided in PLL, it connected with GTX input clock, auxiliary clock and AXI interface clock. 3: The test system has only one global reset clock, which comes form Gemini board reset pin. The GTX reset process is faster than controllers, so we build a 10us delay circuit to make the GTX and controllers reset at the same time.

System test uses Vivado 14.3 ISIM simulation and VCS

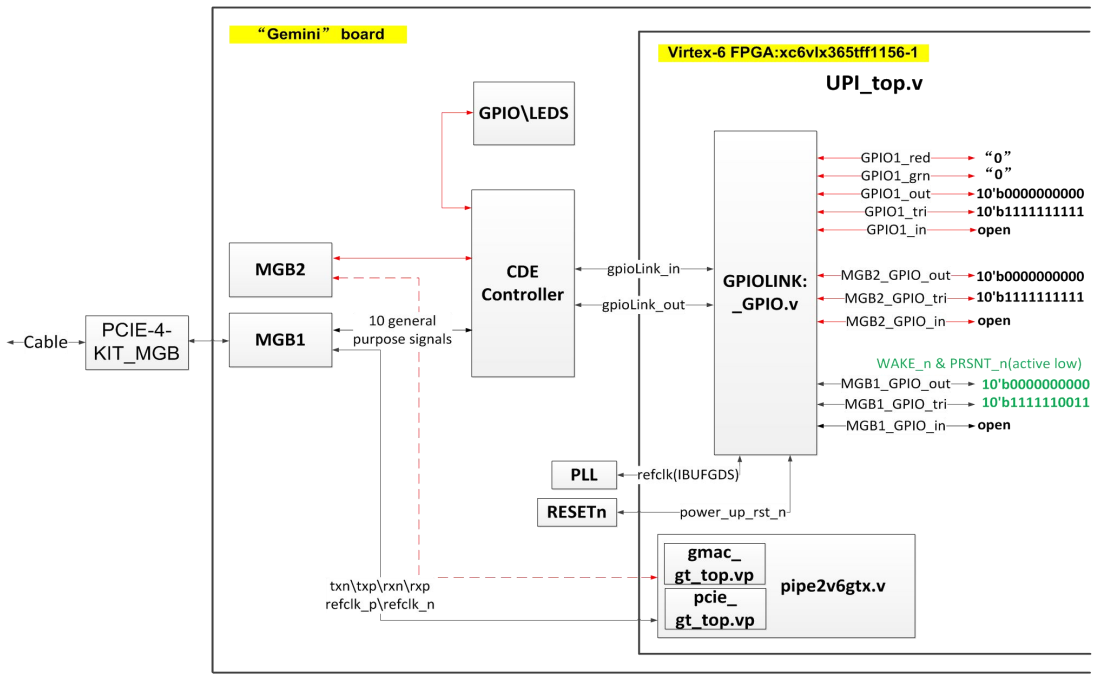


Fig.7. Structure of UPI system test

2014.09-3 simulation tools to verify our design. After functional simulation pass, system test uses Synplify Premier H-2013.03 for RTL implementation and Vivado Chipscope for testing signals observation. The simulation result is shown in Figure 7.

The signal uli_lane_oe_3 available after the other three channels' signals because of the UPI system treat lane four as a special channel. Lane four is used to send control signal and some special communication signals.

Fig.8. UPI system simulation test

Figure 9 shows the FIFO state when UPI system starts to work.

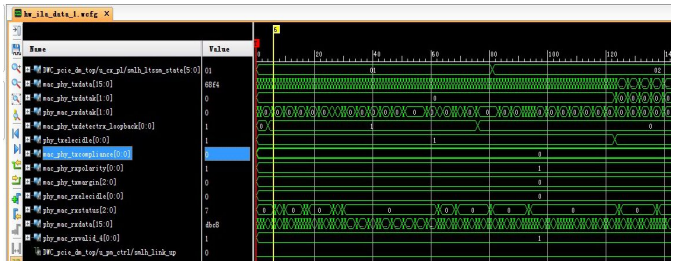


Fig.10. DCM and Interface Convertor states of UPI system

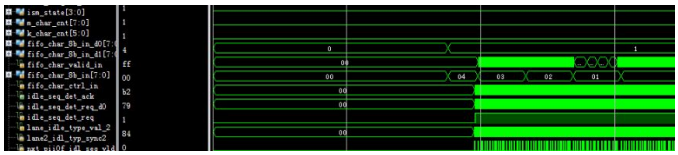


Fig.9. FIFO state of UPI system

Using chipscope to observe the signals come form DCM and Interface convertor, we can clearly see that the current link is in the negotiation state before time of 124us as shown in Figure 10.

After time of 124us, DCM state machine will jump into the next state for exchange link information and link width.

The final placement and layout of the UPI system are shown in Figure 11. Under the control of the PCIe and Ethernet function can be switched by DCM. It uses fewer resources to complete switching function and improves the flexibility of UPI system. This process needs cost some extra area. However, it would increase the system running time significantly for low data throughput or real time applications.

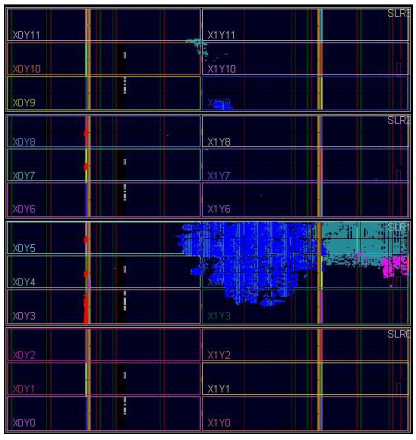


Fig.11. UPI system layout diagram

The design was constrained to the right hand corner of the device where the PCIe and Ethernet blocks resides, as shown in Figure 2. The place with red marked is our UPI wrapper. The maximum clock rate of UPI system is limited only by the design, build tools, and FPGA placing and routing methods.

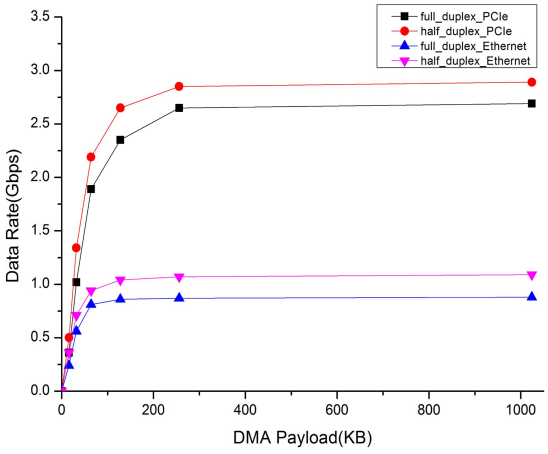


Fig.12. Performance of UPI system

C. *Experimental Results*

Maximum data rates of PCIe and Ethernet are 20Gbps and 4Gbps, respectively. Effective data rate(MB/s) =serial bus clock frequency * 1 Byte(bit/8)* number of ports * encoding format * half-duplex/ full-duplex.

For our design, using X4 lanes, 8/10encoding, full-duplex mode and 2.5GHz(PCIe)/1.25G(GMAC) serial bus clock, effective data rate is 16Gbps/8Gpbs, i.e., 4Gpbs/2Gpbs per lane. System performance can't reach the maximum data rate because of some limits. The highest data rate in test is about 74% of the maximum data rate, calculated from the clock frequency of PLL.

We repeat our test more than 200 times. With the DMA payload sizes keep growing, the performance of UPI also keeps increasing. In DMA payload test, the data rate of both controllers increases as the amount of data increase, as shown in Figure 12. When the DMA payload sizes is larger than 80KB, data rate of Ethernet hold steady at 1.1Gbps while data rate of PCIe still increasing. The data rate of PCIe becomes saturated at 2.8Gbps when payload sizes is larger than 260KB. Ethernet data rate entering saturation more quickly than PCIe because of the maximum packet sizes of Ethernet. Although the data rate of Ethernet is slower than PCIe, the Ethernet physical layer uses 64b/66b encoding, makes Ethernet more effective than PCIe in terms of transmission efficiency.

TABLE III
RESULTS OF DEVICE SWITCHING TEST

	bit error	
Switching time(us)	**full_duplex_PCIe**	**full_duplex_Ethernet**
2.58	100%	100%
1.85	92.6%	93.72%
1.43	46.71%	72.7%
1.04	18.5%	29.35%
0.61	0.037%	0.045%

In Device switching test, we recorded the relationship between bit error and switching time as shown in table 3. UPI can normally work at 2.58us. But as the switching speed accelerated to 1.85us, the bit error increase significantly. When the switching time is less than 500ns, no data entered FPGA B. There are three reasons why the system appears this phenomenon: 1) The switching time is close to the GTX's required reset time, when switching time is less than 500ns, GTX's CPLL is always in the unlocked state, and GTX is unable to complete the transfer task. 2) Some transferring data are stored in FIFO, the data will be flushed after GTX's CPLL reset. In this situation, all data already transmitted are no error. 3) High speed switching causes analog circuits crosstalk especially differential pairs.

When the switching time is more than 2.58us, and the transmission data is greater than 260KB, all transmitted date can be received with no bit errors, except for the write data errors generated by DDR itself.

The resource consumption of UPI system is listed in table 4. We also compare the performance of UPI system with other interconnect I/O technologies as shown in table 4. The results showed that despite the fact that our systems take more resources, UPI system achieves a better flexibility and compatibility. When performance and resources are able to meet the demand of bandwidth between nodes, we made computing nodes of two different protocols compatible.

TABLE IV
RESOURCE CONSUMPTION OVERVIEW

Resource	Our Design in FPGA A	Our Design in FPGA B	UPI system	Resource Available
LUTs	94528(41.5%)	101064(44.4%)	11853(5.2%)	227520
I/O	5(0.4%)	5(0.4%)	N/A	1156
Flip-Flops	65372(14.4%)	84432(18.6%)	8224(1.8%)	455040
BRAM	38(9.1%)	45(10.8%)	6(1.4%)	416

TABLE V
SYSTEM PERFORMANCE OF DIFFERENT INTERCONNECT I/O TECHNOLOGIES

System	Link Rate	Configuration	Link Rate	PCI support	Ethernet support	LUTs
Bluelink[8]	10G, 40G	1x, 4x	10G	No	Yes	2009
Infiniband[9]	40G	LLC QDR 4x	10G	Yes	No	64105
1000 based-X Ethernet MAC	1G	1x	1.25G	Yes	No	11853
PCIe soft IP(stratix IV)	5G	1x Gen2	5G	No	Yes	1805
UPI	11.04G/4.32G	1x, 2x, 3x, 4x	2.76G/1.08G	Yes	Yes	5500

D. Practical Application of UPI

UPI has been used in three practical applications in various institutes. These applications include: MaPU (Mathematics Process Unit) commodity FPGA clusters; HDR (High Dynamic Range) vedio clouding system; MOND (Modified Newtonian Dynamics) hardware accelerator for astronomical data. All these applications require real-time communication with multiple computing nodes. UPI can be used in the application of a good compatibility with the traditional single protocol HPC clusters, as well as adding new UPI-based computing node into computing system. UPI provides a better system compatibility and interface flexibity as shown in table 5.

V. Conclusions and Future Work

A flexible and compatible interconnect interface has been proposed for FPGA-based multi-node communication. We completed a multi-node communication interface with the benefits of high flexibility and good compatibility. We implemented our design on "Gemini" prototype verification board with two Xilinx Virtex-6 FPGAs. The experimental results show that both two bus protocols can be received and transmitted without error when DMA payload size is greater than 260KB (PCIe) and 80KB (Ethernet) and switching time is greater than 2.58us. Through the interface we can easily connect two different HPC clusters. The performance loss caused by traditional bridge equipment is eliminated.

Acknowledgment

This research is supported by the "Strategic Priority Research Program" of the Chinese Academy of Sciences, Grant No.XDA06010402-4.

References

[1] Budruk, Ravi, Don Anderson, and Tom Shanley., "PCI express system architecture," *Addison-Wesley Professional*, 2004.
[2] Intel Corporation's, *PHY Interface for the PCI Express(TM) Architecture*, Specification Version 0.5. pp. 1-15, Aug. 16, 2002.
[3] Gong, Jian, et al., "An efficient and flexible host-fpga pcie communication library," *Field Programmable Logic and Applications (FPL)*, 2014 24th International Conference on. IEEE, 2014.
[4] Koch D, Beckhoff C.,"Hierarchical reconfiguration of FPGAs," *Field Programmable Logic and Applications (FPL)*, 2014 24th International Conference on. IEEE, 2014: 1-8.
[5] Xilinx Inc., *7 Series FPGAs GTX/GTH Transceivers User Guide*, April 22,2013.
[6] Islam, Nusrat S., et al.,"High performance RDMA-based design of HDFS over InfiniBand.," *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis.* IEEE Computer Society Press, 2012.
[7] Shainer, Gilad, et al.,"Maximizing application performance in a multi-core, NUMA-aware compute cluster by multi-level tuning.," *Supercom-puting. Springer Berlin Heidelberg*, 2013.
[8] Theodore Markettos, "A Interconnect for commodity FPGA clusters: standardized or customized?," *Field Programmable Logic and Ap-plications (FPL)*, 2014 24th International Conference on. IEEE, 2014.
[9] TPolybus Systems Corporation, "InfiniBand cores. ," http://www.polybus.com/iblink layer website/ibcores brochure alt.pdf

An Wu received his B.S. degree in 2011 from School of Anhui University, Anhui province, China, and he is currently a Ph.D. student in Department of Physics in University of Science and Technology of China, Anhui, China, under the supervision of Prof. Xi Jin. His current research work is mainly on SoC design technology, VLSI design and FPGA-based Hardware Accelerator design.

Xi Jin received the B.S. degree from University of Science and Technology of China, Anhui, China, and he is currently an associate professor in Department of Physics in University of Science and Technology of China, Anhui, China. His research interests include SOC design technology, VLSI design, computer-aided design methodologies for SoC system integration and FPGA-based Hardware structure design.

Xueliang Du received the Ph.D degree from University of Science and Technology of China, Anhui, China, and he is currently an associate professor in Institute of Automation Chinese Academy of Sciences, Beijing, China. His research interests include High-Performance SoC Design, DSP Design and FPGA-based prototyping.

Shuaizhi Guo received his B.S. degree from University of Science and Technology of China, and he is currently a M.S. student in Department of Physics in University of Science and Technology of China, Anhui, China, under the supervision of Prof. Xi Jin. His current research work is mainly on FPGA-based Hardware Accelerator design.

Accurate Spectral Efficiency Analysis for Non Orthogonal Multiple Access

Pongsatorn Sedtheetorn, Tatcha Chulajata

Department of Electrical Engineering, Faculty of Engineering, Mahidol University

25/25 Phuttamonthon 4 Road, Salaya, Nakornpathom, Thailand

pongsatorn.sed@mahidol.ac.th, tatcha.chu@mahidol.ac.th

***Abstract*— This paper presents theoretical analysis on non-orthogonal multiple access (NOMA) for both downlink and uplink communications. Based on accurate evaluation, we propose closed-form expressions of NOMA downlink and uplink spectral efficiency. In terms of channel modelling, we map the channel conditions with Nakagami and Rayleigh models in which both line-of-sight and non-line-of-sight are included. For uplink transmission, we extend our work to the case of random active users. The random nature is matched to appropriate probability models. These make our presented expressions very comprehensive and benefit us to study the impacts of channel conditions, random number of users, and other key system parameters on the NOMA spectral efficiency.**

***Keyword*— Non orthogonal multiple access, 5G mobile communication, uplink, downlink, spectral efficiency**

I. NTRODUCTION

Recently, the growth of mobile communication is rapidly rising. In the next generation, the entries to mobile networks are not only from human but also from machines. This makes the demand go even greater. It is important noting that radio access technology plays an important role. The current radio access technology, orthogonal frequency multiple access (OFDMA), might not be the best option due to the issues of spectral efficiency and power utilization [1].

As mentioned above, the new radio access technology, namely non-orthogonal multiple access (NOMA), has been introduced [2]-[7]. Based on NOMA, every users share the whole spectrum and each individual is multiplexed to one another via power domain. On the receiver end, the known successive interference cancellation (SIC) method [8] is employed to extract the desired signal from stronger interferences by cancelling (subtracting) them with superposition coding. From the results therein [5], NOMA provides 30% more throughput than the traditional orthogonal multiple access (OMA), which is OFDMA.

In terms of research challenges, there are many aspects to concern. The area on spectral efficiency analysis gains a huge interest, e.g. [3]-[7]. For instance, [4]-[5] focus on the spectral efficiency estimation in which all parameters are however set constantly. The work in [6] is on the analysis of outage probability which is directly related to the spectral efficiency. Others are on the throughput maximization problems such as [3], [7].

Nevertheless, the analysis in the literature still leaves some tasks unfinished. Especially some random-nature parameters such as channel gains and the number of active users are fixed or conditioned owing to the sake of simplicity.

In contrast to the literature, here we present accurate original analysis on NOMA spectral efficiency inclusively on both downlink and uplink transmissions. Furthermore, the impact of random channel gains is mathematically mapped by the famous Nakagami model in which line-of-sight and non-line-of-sight scenarios are taken into account. Also, on the uplink, we model the random number of active users with corresponding probability processes. This makes our work far more practical and outstandingly distinguished from the others.

This paper is organized as follows. Section II illustrates the system model that includes all assumptions and parameters used in this paper. Section III shows the original analysis on NOMA downlink and uplink spectral efficiency. Sequentially, the exact closed forms are presented. Section IV demonstrates the numerical and simulation results. Section V draws the conclusion of this work.

II. SYSTEM MODEL

In this section, the scenario of the future radio access technology, namely NOMA, is explained in terms of both downlink and uplink transmissions. Moreover, all assumptions, key system parameters and performance are defined. Since the transmissions are on wireless channels, the magnitudes of channel gains at a receiver end are inevitably random. Here, the models of channel gains are also classified to cover all possible characteristics of signal propagation.

Manuscript received February 25, 2016. This work is a follow-up of the invited journal to the accepted out-standing conference paper of the 18th International Conference on Advanced Communication Technology (ICACT2016).

P. Sedtheetorn is with the Department of Electrical Engineering, Faculty of Engineering, Mahidol University, Nakornpathom, 73170, Thailand (corresponding author, phone: 66-2889-2225 ext. 6501-3; fax: 66-2889-2225 ext. 6529; e-mail: pongsatorn.sed@mahidol.ac.th).

T. Chulajata is with the Department of Electrical Engineering, Faculty of Engineering, Mahidol University, Nakornpathom, 73170, Thailand (e-mail: tatcha.chu@mahidol.ac.th).

A. Downlink Communication

Figure 1 describes the scenario of NOMA downlink communication. The base station, called eNodeB, serves multiple user equipments (UEs). It is known that NOMA multiplexes individuals in power domain, each of which uses the SIC technique and is able to perfectly decode the signals from the weakest ones.

To demonstrate this, Figure 2 shows the spectrum allocation of NOMA and OMA (e.g. OFDMA). In the figure, the downlink powers of individual UEs are assigned unequally based on their distances from the serving eNodeB. Without the loss of generality, UE1 is assumed to be the closet one with the lowest dedicated power, while UE2, UE3, …, UE N stay further with more dedicated powers. This technique is called rank adaption [2] in which $P_{UE1} < P_{UE2} <\ldots< P_{UEN}$ and $P_{UE1}+P_{UE2}+\ldots+P_{UEN} =P$. Next, the eNodeB sends every UE a downlink signal with the total power P. Then, each UE uses its SIC receiver to extract the desired signal. For example, UE1 uses SIC receiver to filter out stronger signals of UE2, UE3, …, UE N.

In this work, the power spectral of zero-mean additive Gaussian white noise (AWGN) is assumed as N_0. The power gains of UE1, UE2,..,UE N can be defined by $|h_1|^2, |h_2|^2, .., |h_N|^2$. Recall the technique rank adaptation, we have

$$|h_1|^2 / N_{0,1} > |h_2|^2 / N_{0,2} > .. > |h_N|^2 / N_{0,N} \quad (1)$$

where $|h_n|^2 / N_{0,n}$ is defined as signal to noise ratio (SNR) of UE n.

Obviously, the signal power at the receiver end of UE n is

$$S_n = |h_n|^2 P + N_{0,n} \ . \quad (2)$$

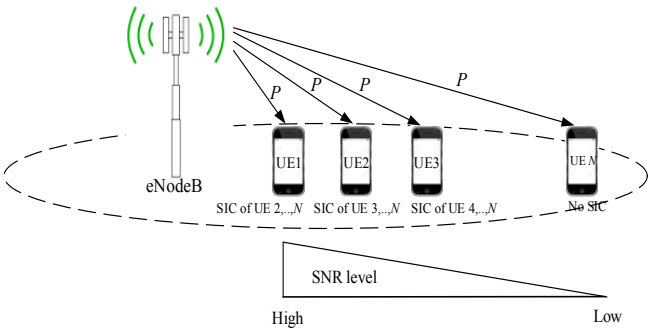


Fig. 1. Downlink NOMA with SIC technique

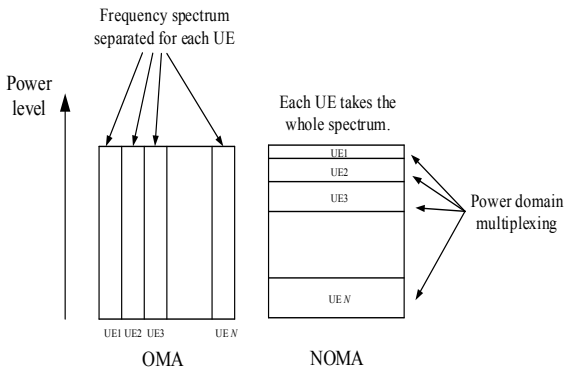


Fig. 2. NOMA vs OMA (OFDMA)

Based on SIC process, UE n, $n \in \{1,2,..,N\}$, can remove the inter-user interference from UE n+1 whose SNR level is smaller, $|h_{n+1}|^2/N_{0,n+1}<|h_n|^2/N_{0,n}$. On the assumption of band-limited waveforms in AWGN channel, the spectral efficiency of UE n can be declared as [4]

$$C_n = \log_2\left(1+\frac{P_n|h_n|^2}{\sum_{i=1}^{n-1}P_i|h_i|^2 + N_{0,n}}\right) \quad (3)$$

which is in bps/Hz and for $n \in \{1,2,..,N\}$. Now, we can define the signal-to-interference-plus-noise ratio (SINR) = $P_n|h_n|^2 \Big/ \sum_{i=1}^{n-1}P_i|h_i|^2 + N_{0,n}$.

B. Uplink Communication

Consider uplink transmissions in Figure 3. Every UE is able to transmit their information through multiple subcarriers, which is called multi-carrier transmissions. Note that the multi-carrier technique, so called carrier aggregation, is preferably used in the future mobile communication in order to increase the overall throughput.

Based on NOMA technique, UEs are allowed to share the same resource simultaneously both in the aspects of frequency spectrum and time. Therefore, the receiver at the eNodeB is required to operate multi-user detecttion (MUD) in order to distinguish signals of individual UEs. One of favorite MUD techniques is SIC in which the desired signal is recovered from the subtraction (cancellation) of interferences.

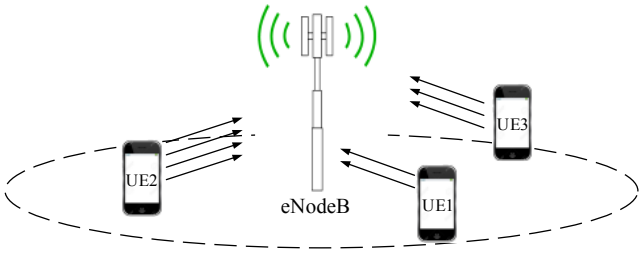


Fig. 3. Uplink multi-carrier NOMA

To this point, the signal received at the eNodeB can be expressed as

$$S_n = \sum_{i=1}^{N}\sum_{l=1}^{L}P_{i,l}|h_{i,l}|^2 + N_0 \quad (4)$$

where L is the maximum number of subcarriers of individual UE. $P_{i,l}$ and $|h_{i,l}|^2$ are the power and the gain of the signal from UE i on the particular subcarrier l.

Similarly, the uplink spectral efficiency of UE n is [3]

$$C_n = \sum_{l=1}^{L}\log_2\left(1+\frac{P_{n,l}|h_{n,l}|^2}{\sum_{i=1,i\neq n}^{N}P_{i,l}|h_{i,l}|^2 + N_0}\right) \quad (5)$$

where the term $\sum_{i=1,i\neq n}^{N}P_{i,l}|h_{i,l}|^2$ acts as the interferences of the desired signal $P_{n,l}|h_{n,l}|^2$ on subcarrier l. Likewise, the uplink SINR = $P_{n,l}|h_{n,l}|^2 \Big/ \sum_{i=1,i\neq n}^{N}P_{i,l}|h_{i,l}|^2 + N_0$.

C. Channel Modelling

Owing to the fact that the power gains are random throughout wireless channel, appropriate probability models are required to complete the analysis. In this work, we apply two famous probability models, namely Rayleigh fading and Nakagami fading.

It is known that Nakagami fading is a comprehensive model that covers both line-of-sight (LOS) and non-line-of-sight (NLOS) scenarios via the index *m* as shown in Table 1 and the mathematical model is illustrated in Section II-C*2*.

Table I.
NAKAGAMI FADING INDEX [9]

Nakagami fading index *m*	Propagation environments
1	NLOS (urban areas)
2-4	obstructed areas
5	suburban areas
6	open plain

In terms of Rayleigh fading, this model is considered as a special case in which a pair of transmitter and receiver stay NLOS. Rayleigh model is regularly used in research work to analyze the extreme case of signal propagation. Furthermore, its mathematical model is expressed by a negative exponential random variable which is quite simple in comparison to the gamma-distributed Nakagami fading.

1) The special case: Rayleigh fading

In the case of NLOS, there is no dominant signal. The received signals scatter from all around directions. A channel gain (voltage gain), $h(t)$, is depicted in terms of imaginary number $h(t) = h_i(t)+\mathrm{j}\ h_j(t)$ in which each component is a Gaussian random variable due to the center limit theory.

Then, the magnitude, $|h(t)|$, stands for Rayleigh process and its square, $|h(t)|^2$, becomes a power gain representing a unit-mean negative exponential random variable with probability density function

$$f_{|h(t)|^2}(z) = e^{-z}. \tag{6}$$

When apply power gains to (3) either or (5), the spectral efficiency becomes a function of multi random variable.

In the literature, the spectral efficiency can be determined by simulating all possible power gain values and then averaging out the result. In some past work e.g. [4]-[5], the gains remain fixed for simplicity.

Alternatively, the average of spectral efficiency can be calculated directly via the multiple integrations of the interferences' density functions. This consumes huge computation time and yields complexity.

In this work, we propose efficient method to accurately calculate the average spectral efficiency of NOMA. Thanks to the properties of exponential distribution, we can use these properties to solve this mathematical difficulty (see Section III).

2) The general case: Nakagami fading

Assume that the signal (channel) gain h_n; models Nakagami fading. The corresponding power gain $|h_n|^2$ becomes a unit-mean gamma random variable with probability density function [10]

$$f_{|h_n|^2}(z) = \frac{z^{m-1}}{\Gamma(m)} m^m e^{-mz} \tag{7}$$

where $0.5 \le m < \infty$ is Nakagami fading index and $\Gamma(m) = (m-1)!$ is the gamma function. When $m = 1$, the probability density function is exponentially distributed which models Rayleigh fading (NLOS fading).

At this point, it is seen that Nakagami fading model covers both NLOS (Rayleigh) and LOS cases. Moreover, the dominant signal (LOS signal) turns to be stronger when *m* goes larger.

III. NOMA SPECTRAL EFFICIENCY ANALYSIS

This section illustrates the analysis of NOMA spectral efficiency both for downlink and uplink. The analysis starts from the special case, Rayleigh fading, to the general case, Nakagami fading, which has more mathematical complexity. As a result, we propose new accurate closed-form expressions of NOMA uplink and downlink spectral efficiency. Furthermore, for uplink, we extend our work to consider the case when the number of active UEs is random.

A. Downlink Analysis

The analysis can be categorized to Rayleigh and Nakagami fading as follows;

1) Rayleigh fading environment

On the condition of Rayleigh fading, the power gains are exponentially distributed random variables. Consider the system model in Section II-A. Now the average spectral efficiency, assumed successful decoding and no error propagation, of UE *n* can be presented as

$$C_{n,avg} = \mathrm{E}\left[\log_2(1+\mathrm{SINR})\right] = \int_0^\infty \log_2(1+z) f_{\mathrm{SINR}}(z) dz. \tag{8}$$

We can see that there is some complexity on the integration of the probability density function of SINR, $f_{\mathrm{SINR}}(z)$. To tackle such the problem, a new efficient method to calculate the spectral efficiency is introduced as below.

Rearrange (8) with the change of logarithmic base, then we have

$$C_{n,avg} = \log_2 e \int_0^\infty \mathrm{P}(\mathrm{SINR} > z) \frac{dz}{1+z} \tag{9}$$

where $f_{\mathrm{SINR}}(z)dz = d\mathrm{P}(\mathrm{SINR} > z)$. Here we use the property of an exponential random variable X, $\mathrm{P}(X > \mu) = e^{-\mu}$, when μ is a constant.

Owing to the fact that $|h_n|^2$ is also exponentially distributed, therefore

$$\mathrm{P}(\mathrm{SINR} > z) = e^{-\frac{z}{P_n}\left(\sum_{i=1}^{n-1} P_i |h_i|^2 + N_{0,n}\right)}. \tag{10}$$

To compute the average spectral efficiency, one needs to find the average of cumulative function $\mathrm{P}(\mathrm{SINR} > z)$. Fortunately, the average value of the cumulative function can be determined by calculating the moment generating function (MGF) of $|h_n|^2$. It is known that the MGF of any exponential

random variable X is $\mathrm{E}\left[e^{-\beta X}\right]=1/(1+\beta)$ for a constant β. Then,

$$\mathrm{E}\left[e^{-\frac{z}{P_n}\left(\sum_{i=1}^{n-1}P_i|h_i|^2+N_{0,n}\right)}\right]=e^{-zN_{0,n}/P_n}\prod_{i=1}^{n-1}\left(\frac{1}{1+(P_i/P_n)z}\right). \quad (11)$$

Replace the MGF derived in (11) into (9). As a result, the closed-form expression of the downlink spectral efficiency of NOMA is

$$C_{n,avg}=\log_2 e\int_0^{\infty}\frac{e^{-zN_{0,n}/P_n}}{1+z}\prod_{i=1}^{n-1}\left(\frac{1}{1+(P_i/P_n)z}\right)dz. \quad (12)$$

Hint that this closed form presents the exact average of the spectral efficiency without any loss of generality in Rayleigh fading environment. Moreover, the closed form can be used in OMA case by simply adding the orthogonal multiplexing factor α.

Denote $\alpha_1, \alpha_2, \ldots, \alpha_N$ by the orthogonal multiplexing factors of UE1, UE2, …, UE N and $\sum_{i=1}^{N}\alpha_i=1$. The power allocation for each UE is identical to one another, $P_1 = P_2 = \ldots = P_N = P$, as a result the spectral efficiency of UE n is

$$C_{n,OMA}=\alpha_n\log_2 e\int_0^{\infty}\frac{e^{-zN_{0,n}/P_n}}{(1+z)^n}dz. \quad (13)$$

2) Nakagami fading environment

Recall (8) with the change of logarithm. Then, we have

$$C_{n,avg}=\log_2 e\left[\int_0^{\infty}\ln(1+z)f_{\mathrm{SINR}}(z)dz\right] \quad (14)$$

Now we cannot use the method as in Section III-1*A* because the density function of SINR is the combination of multiple gamma random variables. Fortunately, we sort this problem by the following *Lemma 1*.

Lemma 1: Let u be a unit-mean gamma random variable with a probability density function $f(u)=\frac{u^{m-1}}{\Gamma(m)}m^m e^{-mu}$ and v be any non-negative random variable and independent from u. Let $z=u/v$ where $u=|h_n|^2$, $v=\frac{1}{P_n|h_n|^2}\left(\sum_{i=1}^{n-1}P_i|h_i|^2+N_0\right)$, and

$$\mathrm{E}\left[\ln(1+z)|v\right]=\int_0^{\infty}\frac{1}{z}\left[1-\frac{1}{(1+z)^m}\right]\mathrm{MGF}(mz)dz. \quad (15)$$

Proof: we know that $\ln(1+z)=z\,{}_2\mathrm{F}_1(1;1;2;-z)$ where ${}_2\mathrm{F}_1(a;b;c;z)$ is a Gauss hyper geometric function [11]. Next we get

$$\begin{aligned}&\frac{1}{(m-1)!}\frac{d^m}{dz^m}z^{m-1}\ln(1+z)\\&=\frac{1}{(m-1)!}\frac{d^m}{dz^m}z^m\,{}_2\mathrm{F}_1(1;1;2;-z)\\&=m\,{}_2\mathrm{F}_1(1+m;1;2;-z)\\&=m\int_0^{\infty}(1+tz)^{-(m+1)}dt=\frac{1}{z}-\frac{1}{z(1+z)^m}.\end{aligned} \quad (16)$$

Also, the MGF can be solved in a simpler form,

$$\mathrm{MGF}(mz)=e^{-zmN_0/P_n}\prod_{i=1}^{n-1}\frac{1}{\left(1+\frac{P_i}{P_n}z\right)}. \quad (17)$$

Replace $\mathrm{MGF}(mz)$ in (15) and (14) respectively. As a result, the average spectral efficiency of downlink NOMA in Nakagami fading is

$$C_{n,avg}=\log_2 e\left[\int_0^{\infty}\frac{e^{-zmN_{0,n}/P_n}}{z}\left[1-\frac{1}{(1+z)^m}\right]\prod_{i=1}^{n-1}\frac{1}{\left(1+\frac{P_i}{P_n}z\right)}dz\right]. \quad (18)$$

B. Uplink Analysis

Likewise, the closed-form spectral efficiency of uplink can be derived by the same procedures as those of downlink. However, another issue on the uplink is the randomness of active UEs. In practice, the number of active UEs (during voice either or data transmission) is unknown to the eNodeB. Here we investigate the effect of random number of UEs and extend our closed forms to include this case.

1) Fixed number of users

Recall (5) and the procedure to achieve (15). We derive the closed form of uplink NOMA spectral efficiency in Rayleigh fading as

$$C_{n,avg}=\log_2 e\sum_{l=1}^{L}\left\{\int_0^{\infty}\frac{e^{-zN_0/P_{n,l}}}{1+z}\prod_{i=1,i\neq n}^{N}\left(\frac{1}{1+(P_{i,l}/P_{n,l})z}\right)dz\right\}. \quad (19)$$

Similarly, recall the same procedure to accomplish (18). We have the uplink spectral efficiency in Nakagami fading as

$$\begin{aligned}C_{n,avg}=\log_2 e\sum_{l=1}^{L}\Bigg\{&\int_0^{\infty}\frac{e^{-zmN_0/P_{n,l}}}{z}\left[1-\frac{1}{(1+z)^m}\right]\\&\prod_{i=1,i\neq n}^{N}\left(\frac{1}{1+(P_{i,l}/P_{n,l})z}\right)^{-1}dz\Bigg\}.\end{aligned} \quad (20)$$

Let all UEs be active simultaneously, then the total spectral efficiency is as $C_{tot}=NC_n$. In practice, the number of active UEs is however random and distributed from 0, 1, 2, …, N. In the next section, we therefore sort this problem and put the effect of random UEs in the proposed expressions.

2) Random number of users

In this paper, we can model the randomness of active UEs by Poisson and binomial distributions. The Poisson model seems reasonable when the total number of both active and inactive UEs is unknown at the eNodeB. Then the amount of active UEs is estimated by statistically average value. For binomial model, the total number of UEs is determined and the number of active UEs is defined by a certain on-off (active-inactive) probability. These assumptions (Poisson and binomial) could be available in practice.

a) Poisson distribution

In terms of Poisson process, the number of active users is approximated by the average value λ. Then, the probability mass function (pmf) can be written as [10]

$$\Pr(N=k)=\frac{\lambda^k}{k!}e^{-\lambda} \quad (21)$$

where k is the number of active UEs at an instance. Hint that, when the number of interfering UEs is random, we need to recalculate $C_{n,avg}$. For example, recall (19) and assume all transmitted uplink powers are identical. Without the loss of generality, the spectral efficiency becomes

$$\mathrm{E}\left[e^{\frac{-zm}{P_{n,l}}\left(\sum_{i=1,i\neq n}^{N} P_{i,l}\left|h_{i,l}\right|^2+N_0\right)}\right]=e^{-\lambda}\left(\frac{1}{1+z}\right)^{N-1}. \quad (22)$$

Lemma 2: Let $V(\cdot)$ be any function and Y be a Poisson random variable with parameter (λ). Then, we have

$$\mathrm{E}\left[V^{Y}(\cdot)\right]=e^{-\lambda[V(\cdot)-1]}. \quad (23)$$

Proof: Let t be a constant. The moment generating function of Y is $\mathrm{E}\left[e^{tY}\right]$,

$$\mathrm{MGF}(Y)=\mathrm{E}\left[e^{tY}\right]=\sum_{k=0}^{\infty}e^{tk}\frac{\lambda^k}{k!}=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\left(e^{t}\lambda\right)^k}{k!}. \quad (24)$$

We know that the summation of the pmf is equal to 1; i.e. $\sum_{k=0}^{\infty}e^{-\lambda}\lambda^k/k!=1$, then $\sum_{k=0}^{\infty}\lambda^k/k!=e^{\lambda}$. Apply this to MGF($Y$), thus it becomes

$$\mathrm{MGF}(Y)=e^{-\lambda}e^{e^{t}\lambda}=e^{\lambda(e^{t}-1)}. \quad (25)$$

Replace e^t with function $V(\cdot)$. Finally we have the proof of *Lemma 2*.

Apply *Lemma 2* to (19). Then, we have the spectral efficiency of individual UE. As a result, the total spectral efficiency becomes

$$C_{tot}=\lambda C_n. \quad (26)$$

Note that we can apply *Lemma 2* to (20) for the spectral efficiency in Nakagami fading as well.

b) Binomial distribution

Recall (22). Now N is binomially distributed. Let us declare *Lemma 3*.

Lemma 3: Let $V(\cdot)$ be any function and Y be a binomial random variable with parameters (N, p). Then, we have

$$\mathrm{E}\left[V^{Y}(\cdot)\right]=\left[1-p+pV(\cdot)\right]^{N} \quad (27)$$

Proof: if Y is a binomial random variable with parameters (N, p). The MGF of Y is $\mathrm{E}\left[e^{tY}\right]$ in which t is a constant. Then,

$$\begin{aligned}\mathrm{E}\left[e^{tY}\right]&=\sum_{k=0}^{N}e^{tk}\binom{N}{k}p^k(1-p)^{N-k}\\&=\sum_{k=0}^{N}\binom{N}{k}(pe^{t})^k(1-p)^{N-k}=\left[1-p+pe^{t}\right]^{N}.\end{aligned} \quad (28)$$

Replace e^t with function $V(\cdot)$, thus we yield the same result as in (27).

Apply (27) to (19). The exact average uplink spectral efficiency is

$$C_{n,avg}=\log_2 e\sum_{l=1}^{L}\left\{\int_0^{\infty}\frac{e^{-zN_0/P_{n,l}}}{1+z}\left[1-p+p\left(\frac{1}{1+z}\right)\right]^{N-1}dz\right\}. \quad (29)$$

As a result, the overall spectral efficiency is

$$C_{tot}=pNC_n \quad (30)$$

where pN is the average value of a binomial random variable with parameter (N, p). Note that the spectral efficiency in the case of Nakagami fading can be derived by the same manner.

IV. Numerical Results and Discussions

In this section, selected numerical results and simulation results are shown. The results are categorized to uplink and downlink aspects each which represents different scenarios of NOMA spectral efficiency versus key system parameters. Furthermore, we also validate our numerical results, proposed from our expressions, with the simulation. In terms of simulation results, we apply Monte Carlo simulation in which the random power gains are generated and the spectral efficiency is computed by the original Shannon formula (see equation (3) and (5) for downlink and uplink respectively). Then, we repeat the iterations over 2,000,000 times to make sure the results reliable.

A. Downlink Communication

Consider a single-cell environment with three UEs, namely UE1, UE2, and UE3, respectively. UE1 is the nearest one to the eNodeB whereas UE2 stays further away from the eNodeB and UE3 is the furthest. Then, the power allocations are P_1=1/6, P_2=1/3, P_3=1/2 for UE1, UE2, and UE3, respectively. Let SNR = P/N_0 where P= P_1+ P_2+ P_3. Assume that the fading index m=1 represents non-line-of-sight propagation which models Rayleigh fading.

In Figure 4, our numerical results, generated from (12), is positively matched to those of the simulation. This validates the accuracy of our proposed expression. In the figure, the expression is used to evaluate the NOMA spectral efficiency of each UE against the SNR in dB. Obviously, the spectral efficiency of UE1 (the nearest one) is higher than others because it has the highest rank adaptation (highest SINR) which supports the principle of NOMA with SIC receivers.

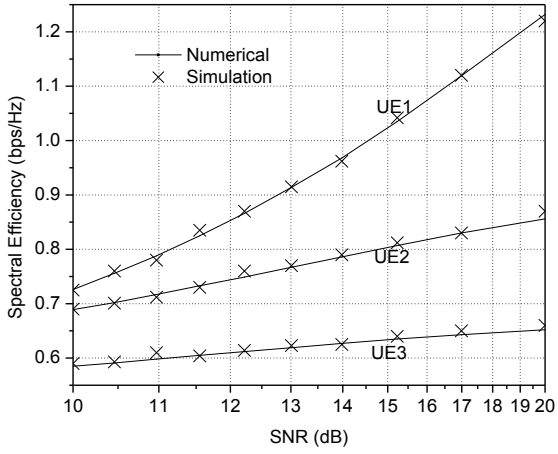


Fig. 4. NOMA spectral efficiency: validation of the proposed expression (numerical result) with the simulation

In addition, we can extend our work to compute the spectral efficiency of OMA (OFDMA) (see equation (13)). From Figure 5, the overall spectral efficiency of NOMA is up to 30% higher than those of OMA which is identical to the results in [5].

Figure 5 demonstrates four different power allocation plans, i.e. A, B, C, and D with various power proportions for individual UEs. Note that the spectral efficiency plotted in this figure is the summed value, i.e. C= C_1+ C_2+ C_3. It can be seen that the power proportion of far UE should be greater to gain better overall spectral efficiency (plan A). With this allocation plan, the spectral efficiency however rapidly drops when SNR goes lower than 17 dB. Thus, the power allocation plan B seems the most optimal solution in this scenario.

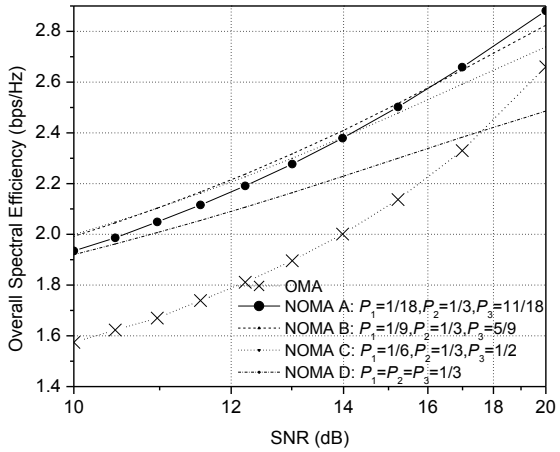


Fig. 5. NOMA spectral efficiency with different power allocation plans (plan A, B, C, and D) versus OMA (OFDMA)

Next, the evaluation of NOMA spectral efficiency in Nakagami fading is taken into account (see equation (18)). Figure 6 illustrates the impact of channel conditions on the spectral efficiency. Here the channel conditions are represented by the integer fading index *m*. It is known that, when *m*=1, the channel acts as Rayleigh model in which the line-of-sight between transmitters and receivers disappears. Otherwise, when m is larger, the dominant line-of-sight becomes more obvious and the spectral efficiency increases.

From Figure 6, it is worth saying that the impact of fading index *m* on the spectral efficiency is greater than that of SNR. We can observe that, when *m* =1,2 (highly obstructed channel), the impact of SNR, considered as background noise, is unobvious. On the other hand, when the channel is clear (high *m*), SNR plays a role in the degradation of spectral efficiency.

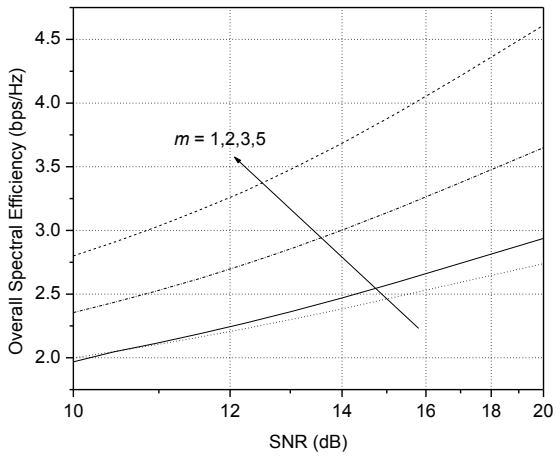


Fig. 6. NOMA spectral efficiency with different fading indices

B. Uplink Communication

In this section, we focus on uplink communication. Assume that there are 5 active UEs, each of which occupies 3 subcarriers, in the cell and their power transmissions are in the same level. Hint that signal-to-noise ratio (*SNR*) is defined as SNR = $P_{n,l}/N_0$ in decibel (dB).

Figure 7 illustrates the impact of channel conditions on the spectral efficiency. It is known that, when $m = 1$, the channel acts as Rayleigh model in which there is no LOS between the eNodeB and its UEs. On the other hand, when *m* is larger, the dominant line-of-sight becomes more obvious and the spectral efficiency increases.

Again, the solid lines, representing numerical results, are matched to the cross symbols of the simulation. This can confirm the validation of our proposed expression in (20).

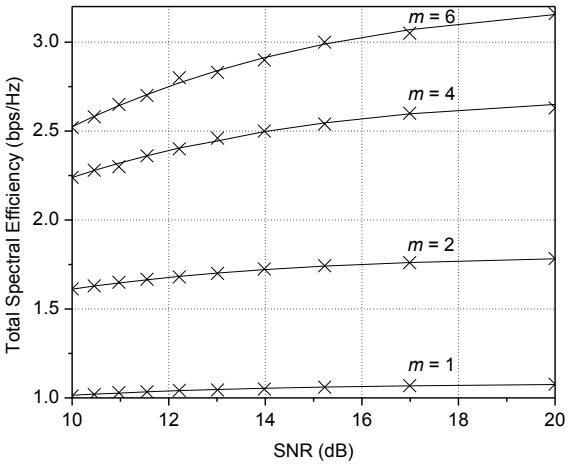


Fig. 7. Total uplink spectral efficiency at different fading index

In Figure 8, we plot the uplink spectral efficiency (equation (26)) versus SNR (*m*=1) in the case that number of active UEs are random and Poisson distributed. Assume that all UEs use 3 subcarriers to convey their information (*L*=3). λ varies from 2, 3, 5 and 10. It is seen that increasing number of active UEs reflects on interferences to one another and the decrease of the spectral efficiency.

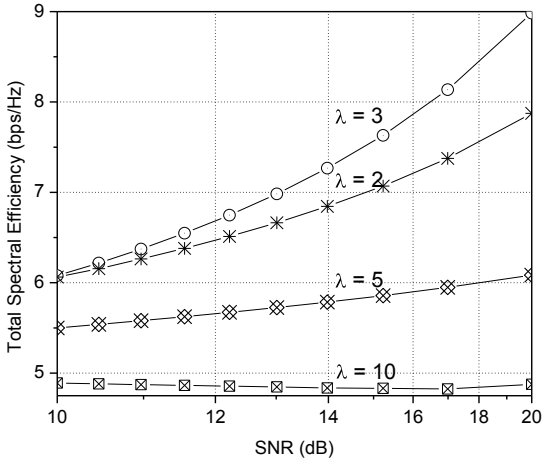


Fig. 8. Total uplink spectral efficiency at varying arrival rate

In Figure 9, the total spectral efficiency versus SNR with varying active probability in (30) is concerned. Here the number of active UEs is binomially distributed. The total number of UEs is *N*=10. All UEs use 3 subcarriers to convey their information (*L*=3). In the figure, the active probability (*p*) varies from 0.1, 0.3, 0.5, 0.7, and 0.9.

It is seen that high value of active probability activates UEs to transmit their signals and thus interfere one another. This leads to the drop of the overall spectral efficiency. Moreover, we find that the appropriate value is around 0.3 or

3 active users at an instance which is equal to the optimal Poisson arrival rate (λ=3) in Figure 8.

To have more demonstration on this issue, let see Figure 10. SNR is set as 10 dB and all UEs use 3 subcarriers to convey their information while the total number of UEs (N) varies from 5, 7, 15, and 20. From the figure, we can get the optimal condition when the total number of UEs is around 3.

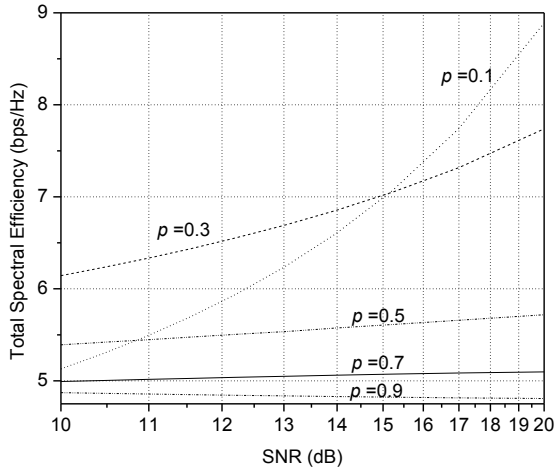


Fig. 9. Total uplink spectral efficiency at different active probabilities

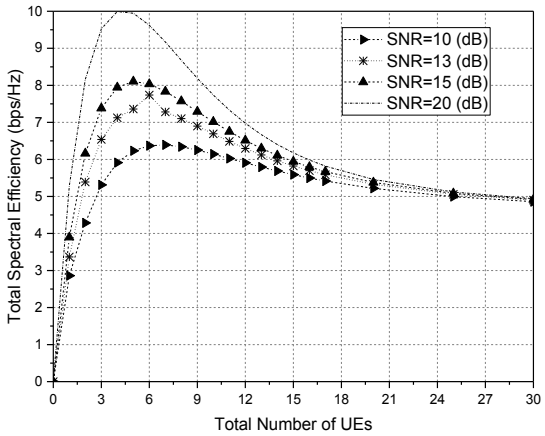


Fig. 10. Total uplink spectral efficiency at different values of SNR

V. Conclusion

This paper has presented original analysis on NOMA spectral efficiency for both downlink and uplink. This has introduced new exact closed forms of NOMA downlink and uplink spectral efficiency. The closed forms also reflect the channel condition via Nakagami fading index. Furthermore, the effect of random users on uplink is taken into account and characterized with different probability models. This benefits us to accurately investigate the impacts of channel conditions, random active users, and key system parameters on the NOMA spectral efficiency.

From the results, we have found that the impact of channel blockage on the spectral efficiency is far more obvious than that of the background noise (represented by SNR). Furthermore, especially on the uplink, the number of active users directly reflects the amount of interferences. When the arrival rate (for Poisson model) or active probability (for binomial) is greater, the overall spectral efficiency decreases. However, there is an optimal number of active users to achieve the highest spectral efficiency at a certain provided bandwidth. In this paper, the optimal point of binomial distributed model is equal to that of Poisson distributed. This also reasonably confirms and validates our proposed expressions.

References

[1] Jeffrey G. Andrews *et al*., "What will 5G be?," *IEEE Journal on Selected Areas in Commun.*, vol. 32, no. 6, pp. 1065-1082, June 2014.

[2] Docomo 5G white paper, "5G radio access: requirements, concept and technologies," NTT Docomo Inc., 2014.

[3] Mohammed Al-Imari *et al.*, "Uplink non-orthogonal multiple access for 5G wireless networks," in *IEEE Proceeding of ISWCS*, vol. 1, August 2014, pp. 781-785.

[4] Yuya Saito *et al*., "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *IEEE Proceeding of VTC Spring*, vol. 1, June 2013, pp. 1-5.

[5] Anass Benjebbour *et al*., "Concept and practical considerations of non-orthogonal multiple access (NOMA) for future radio access," in *IEEE Proceeding of ISPACS*, vol. 1, November 2013, pp. 770-774.

[6] Zhiquo Ding *et al*., "On the performance of non-orthgonal multiple access in 5G systems with randomly deployed users," *IEEE Trans. on Signal Processing Lett.*, vol. 21, no. 12, pp. 1501-1505, December 2014.

[7] Stelios Timotheou *et al*., "Fairness for non-orthogonal multiple access in 5G systems," *IEEE Trans. on Signal Processing Lett*., vol. 22, no. 10, pp. 1647-1651, October 2015.

[8] Mazen O. Hasna *et al*., "Performance analysis of mobile cellular systems with successive co-channel interference cancellation," *IEEE Trans. on Wireless Commun*., vol. 2, issue 1, pp. 29-40, February 2003.

[9] John G. Proakis *et al.*, *Digital Communications*, 5th edition, Mc-GrawHill, 2008.

[10] Sheldon M. Ross, Introduction to probability models, 7th edition, Harcourt Academic Press, 2000.

[11] S.M. Abramowitz, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, U.S. Department of Commerce, 1972.

Pongsatorn Sedtheetorn (M'03) received the B.Eng. and M.Eng. degrees from Chulalongkorn University, Thailand, in 1998 and 2001, and the Ph.D. degree from the University of Manchester, United Kindom, in 2007. He is currently an Associate Professor with the Department of Electrical Engineering, Mahidol University, Thailand. His research interests are in the areas of wireless communications, information theory, as well as enterprise architecture.

Tatcha Chulajata (M'97) received the B.Eng. from Kasetsart University, Thailand, in 1992. He received the M.S and the Ph.D. degrees from Wichita State University, USA, in 1996 and 2003, respectively. He is currently a Senior Lecturer with the Department of Electrical Engineering, Mahidol University, Thailand. His research interests are in the areas of wireless communications, communication network, and enterprise architecture.

Your Neighbors Are My Spies: Location and other Privacy Concerns in GLBT-focused Location-based Dating Applications

Nguyen Phong HOANG, Yasuhito ASANO, Masatoshi YOSHIKAWA

Department of Social Informatics, Graduate School of Informatics, Kyoto University

Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan

hoang.nguyenphong.jp@ieee.org, asano@i.kyoto-u.ac.jp, yoshikawa@i.kyoto-u.ac.jp

***Abstract*—Trilateration is one of the well-known threat models to the user's location privacy in location-based apps; especially those contain highly sensitive information such as dating apps. The threat model mainly bases on the publicly shown distance from a targeted victim to the adversary to pinpoint the victim's location. As a countermeasure, most of location-based apps have already implemented the "hide distance" function, or added noise to the publicly shown distance in order to protect their user's location privacy. The effectiveness of such approaches however is still questionable. Therefore, in this paper, we investigate how the popular location-based dating apps are currently protecting their user's privacy by testing three popular GLBT-focused apps: Grindr, Jack'd, and Hornet. We found that Jack'd has the most privacy issues among the three apps. As one of our findings, we also show how the adversary can still figure out the location of a targeted victim even when the "show distance" function is disabled in Grindr. Without using sophisticated hacking techniques, our proposed model (*called colluding-trilateration*) is still very effective and efficient at locating the targeted victim, and of course in a so-called "legal" manner, because we only utilize the information that can be obtained just as same as any other ordinary user. In case of Hornet, although it has adopted location obfuscation in its system, we were not only able to discover its noise-adding pattern by conducting empirical analysis, but also able to apply the colluding trilateration used in Grindr to locate the targeted victim regardless of the location obfuscation. Our study thus raises an urgent alarm to the users of those location-based apps in general and GLBT-focused dating apps in particular about their privacy. Finally, the paper concludes by suggesting some possible solutions from the viewpoints of both the LBS provider and the user considering the implementation cost and the trade-off of utility.**

***Keyword*—Location-based Application, GLBT-focused Applications, Location Privacy, User Privacy, Trilateration, Colluding-trilateration, Grindr, Jack'd, Hornet**

Manuscript received on March 11th, 2016. This work was supported by JSPS KAKENHI Grant Number 15K00423 and the Kayamori Foundation of Informational Science Advancement. The paper is a follow-up of [21], which was presented at the 18th IEEE International Conference on Advanced Communication Technology, and received the Outstanding Paper Award.

Nguyen Phong HOANG is currently a graduate student at the Department of Social Informatics, Graduate School of Informatics, Kyoto University, Japan. (corresponding author: +81-75-753-5375, fax: +81-75-753-5375, e-mail: hoang.nguyenphong.jp@ieee.org)

Yasuhito ASANO is an associate professor at the Graduate School of Informatics, Kyoto University (e-mail: asano@i.kyoto-u.ac.jp).

Masatoshi YOSHIKAWA is a professor at the Graduate School of Informatics, Kyoto University (e-mail: yoshikawa@i.kyoto-u.ac.jp).

I. INTRODUCTION

NOWADAYS, thanks to the advancement of the Global Positioning System (GPS), most of the smart phones have a built-in GPS receiver, which assists to estimate the location information with accuracy up to just a few meters. Taking this advantage, location-based applications (*aka: location-based services or LBS*) are getting more dominant in the smart phone application market. Just a decade ago, one still had to use paper map or ask for direction when going to an unfamiliar area; while young people were surfing around online chat rooms to look for friends at that time. However, the introduction of LBS has changed our lifestyle and the way that people interact with each other thanks to its undeniable convenience. For instance, one can easily find the nearest restaurant, convenience store or shopping mall by using application like Google Map; or hang out with friends by using application like Find My Friends, etc.

A. Privacy in General

Nevertheless, in the era of Information and Communications Technology, along with censorship and massive surveillance in cyberspace, the problem of information leakage has also become more and more severe. Tim Cook, the CEO of Apple Inc., used to say at the White House Cyber Security Summit in early 2015 that: "Privacy is a matter of life and death" [1]. As people increasingly keep more sensitive personal information in their phone, big agencies and companies like Apple have been working hard to provide the best protection to their customer's private information. However, an absolute privacy and a completely perfect countermeasure to prevent future data breaches still remain as headache matters. According to the Tenth Annual Cost of Data Breach Study published by IBM in 2015, the average consolidated total cost of a data breach is $3.8 million, increasing 23% since 2013. The report also points out that the cost incurred for each lost or stolen record that has sensitive and confidential information increased 6% from a consolidated average of $145 to $154 [2].

Among personally identifiable information, location is considered as one of the most essential factors since the leak of location information can consequently lead to the disclosure of other sensitive private information such as occupations, hobbies, daily routines, and social relationships [3]. In spite of many attack techniques [4], [5] that have been

studied by the research community since then, the protection of location privacy from both LBS provider and user has not been sufficiently and appropriately taken into account. Thus, in this study, we investigate the current status of location privacy preserving in popular GLBT-focused dating applications to have a clearer view on the issue, and observe how it is being protected in the real-life practice under both already-known threat (*i.e. trilateration*) and its enhanced version (*i.e. colluding trilateration*) proposed by our group.

B. Privacy Concerns in GLBT-focused Applications

First of all, it is important to emphasize that it is not because of hatred or discrimination that makes us opt for investigating GLBT-focused applications like Grindr, Jack'd and Hornet. But, because of their popularity, possession of highly-sensitive information, and the huge number of users[1] that make these applications highly vulnerable to cyber-attack like the case of Ashley Madison [6]. In addition, it is also because GLBT-focused dating applications like Grindr, Jack'd, and Hornet are location-sensitive, and their users depend on the publicly shown distance information to look for nearby people to meet up right away for hookup (*most of the time*), thus potentially exposed to the risk of being located.

As stated in [7], there are still many Islamic nations where homosexuality carries the death penalty. Most recently, there were several gay men in Syria lured by ISIS terrorists to go out on dates, and later executed publicly by stoning as reported in [8]. Even in those regions like North America and Western Countries, which are thought to be more open-minded, the GLBT community is still not widely accepted. More or less, people belong to the GLBT community are still facing the problem of being attacked, harassed or discriminated [9]. Such cases show that protecting privacy of the user of GLBT-focused application is a nontrivial task, and should not be neglected by the LBS providers. Because the location information together with other information such as height, weight, age, and hobby can be used to accurately disclose the targeted individuals. Later, the compromised information from those victims such as occupation, address, or frequently visiting places, daily routines and social relationships can be used to intimidate for money, or even lead to physical harassment. At this point, it is understandable why Tim Cook says: "Privacy is a matter of life and death" since he also came out as a member of the GLBT community in October 2014 [10].

C. Organization

The rest of this paper is organized as follows. We will introduce our experimental environment in the last part of this section. In Section II, Jack'd, Grindr and Hornet are investigated in terms of location privacy. By employing our proposed colluding-trilateration, we will demonstrate how the user's location still can be accurately discovered even when countermeasures like location anonymization and location obfuscation have been implemented. In Section III, other privacy concerns are discussed with real life experiments. In this section, we will also introduce a side-channel attack fashion that can be conducted due to the current design of Jack'd. Finally, from the viewpoints of both LBS provider and user, we then give some possible solutions, and wrap up the paper in Section IV.

D. Experiment Setup

The trilateration threat model actually can be conducted in a physical way that the adversary carries his device around to three different places and notes down the distances shown from his position to the victim. However, in order to have an easily manageable experimental environment, we employ three virtual machines that host Android OS to play the role of adversaries. Each machine is then set to be in positions around our institute as follows:

- Victim is an account run on a real iPhone 5, locates at Science Frontier Laboratory, Kyoto University with coordinates (35.02350485, 135.77687703).
- Adversary A1 is located at Demachi-yanagi Station with coordinates (35.03051251, 135.77327415).
- Adversary A2 is located at Heian Shrine with coordinates (35.01598257, 135.78242585).
- Adversary A3 is located at Kyoto Imperial Palace with coordinates (35.02258561, 135.76493382).

Each Android machine is then equipped with Fake-GPS[2] so that their positions can be freely set to any corner of the world. At the time of writing this paper, we did our experiments with Grindr (version 2.2.8), Jack'd (version 3.3.2), and Hornet (version 2.7.1). Next, to capture packets in Subsection III.A, we set up a proxy machine, and use Microsoft Network Monitor (version 3.4) to monitor network traffic passing through that proxy machine. All of the maps used in this study are sketched using a map tool available at: http://obeattie.github.io/gmaps-radius/.

II. LOCATION PRIVACY CONCERN

To initially test whether an application adopts location obfuscation to obscure the publicly shown distance of its user or not, we move around two accounts run in the virtual environment mentioned in Subsection I.D. The distance shown on each account is then recorded and compared with the real distance. From some preliminary results, we found that Grindr and Jack'd (*in default setting*) do not adopt any location randomization, but show the exact physical distance of the user. As a consequence, the real location of the user is vulnerable to the trilateration threat model, which will be discussed in more detail in Subsection II.A. To prevent the risk of being located for its user, Grindr has already implemented a function which allows the user to hide the distance from being viewed by other users, while Jack'd has not implemented any effective countermeasure to alleviate this risk. Nonetheless, in Subsection II.B, by deploying our proposed colluding-trilateration method, we will demonstrate that disabling the "show distance" function still cannot effectively mitigate the risk. In contrast with Grindr and Jack'd, Hornet seems to be better in protecting its user's location privacy by adopting location obfuscation in its system. As a result, we always get the distance shown on the

[1] According to [22], both Grindr and Jack'd currently have more than 5 million active users, while there are more than 4 million active users in Hornet. To examine that, at the time of writing this paper, we tried inputting the terms such as "gay dating" or "gay hookup" in to Appcrawlr, which is a semantic mobile application discovery powered by Softonic. The search engine indeed returned Grindr, Jack'd and Hornet on the top of the result list sorted in the order of number of downloads.

[2] https://play.google.com/store/apps/details?id=com.lexa.fakegps

application different from the real distance. However, in Subsection II.C, we will show how our colluding-trilateration model can still be applied to precisely locate Hornet user regardless of whether location anonymization and location obfuscation are enabled at a same time in Hornet.

A. *Trilateration Model*

As far as we are aware, the trilateration threat model (*aka: triangulation*) is said to be first reported to Grindr in 2014 [11], and discussed in recent studies [4] and [5]. The main idea of this attack model bases on the distance from the user to the adversary, which is publicly shown to other users. With privilege no more than an ordinary user, an adversary just needs to move around the victim to three different places. Distances from the victim to the adversary at those three positions are then used to pinpoint the exact location of the victim. As shown in Figure 1, Grindr and Jack'd users, who keep the default setting, are facing a high risk of being located since the adversary can obtain an accurate location up to the victim's building as highlighted in the red rectangle. With this threat model, we could not locate the victim in Hornet because the real distance is obscured, and the publicly shown distance is changed to new value every time we re-query it. Let us revisit Hornet in Subsection II.C.

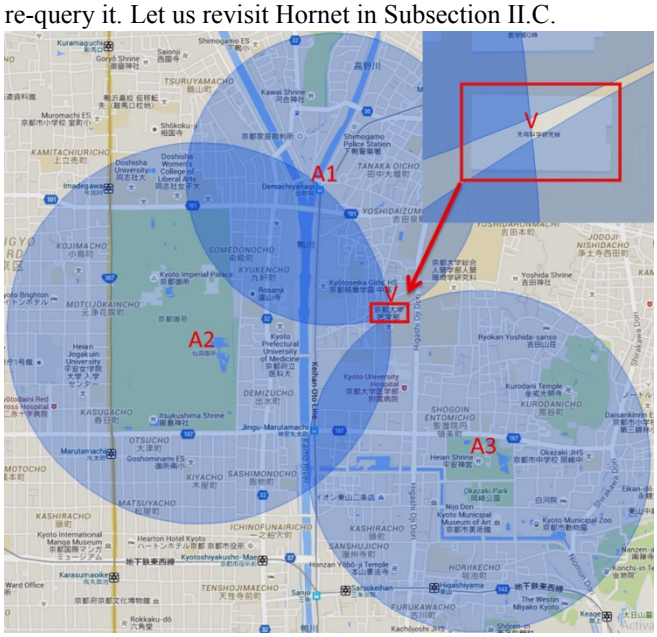


Fig. 1 Testing Trilateration Threat Model in Grindr and Jack'd.

From the geometry point of view, the location of the victim is nothing else but the coordinates of V, which is the solution (x, y) of a system of simultaneous circle equations.

$$\begin{cases}(x - x_{A1})^2 + (y - y_{A1})^2 = D1^2 \\ (x - x_{A2})^2 + (y - y_{A2})^2 = D2^2 \\ (x - x_{A3})^2 + (y - y_{A3})^2 = D3^2\end{cases}$$

Where:

- (x_{A1}, y_{A1}), (x_{A2}, y_{A2}) and (x_{A3}, y_{A3}) are latitudes and longitudes of the adversary at three different positions A1, A2, A3 respectively.
- D1, D2, and D3 are distances from V to A1, A2, and A3 respectively.

In response to this type of threat, Grindr has adopted a function in which the user can opt to hide the distance since August 2014 [12]. Thus, the trilateration model is no longer able to locate those users who already disabled the "show distance" function. As we revisited [11] at the time of writing this paper, the map is no longer able to pinpoint Grindr users as shown in Figure 2.

Fig. 2 Previous Grindr's flaw had been fixed.

For Jack'd, it does not adopt the location anonymization policy to protect its user. Instead, it creates a function which allows its user to adjust the accuracy of the distance to three levels: close, near, and far (*in iOS*); or street, neighborhood, and city (*in Android*) as shown in Figure 3.

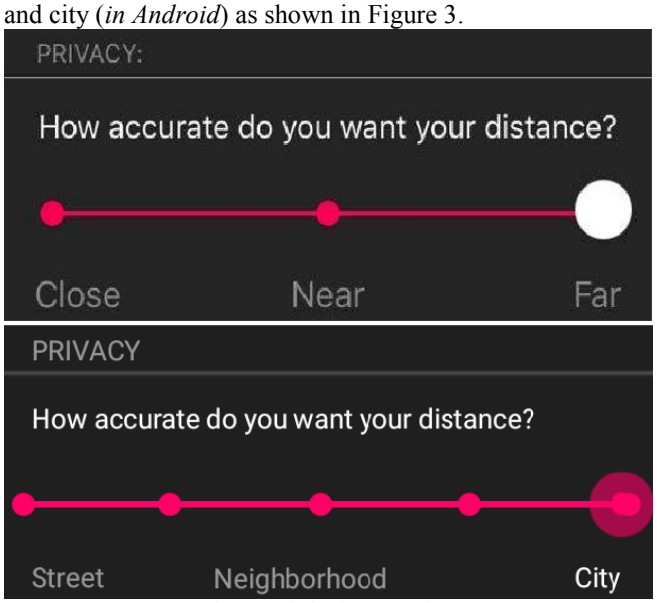


Fig. 3 Jack'd Privacy Setting.

Notwithstanding these setting options, the publicly shown distance between two of our fake accounts does not change even when we restart the application to load the new privacy setting. As a result, this new function of Jack'd does not guarantee the location privacy of its user at all.

B. *Your Neighbors are My Spies – Colluding Trilateration*

Despite of the fact that the best solution to protect the location information is not to publish it; in this part, as a key point of this paper, we will illustrate an enhanced version of the trilateration threat model that current approach like location anonymization implemented by disabling the "show distance" function still cannot effectively counter to. The primary factor in the success of this threat model bases on the way that Grindr arranges its users on the screen. Perhaps, in order to provide a high utility for the application, users are displayed left-to-right and top-to-down in an ascending order of their distances regardless of whether they have already

disabled the "show distance" function or not. By exploiting this fact, the two neighbors appear just before and just behind the victim on the application's screen unintentionally become the upper and lower bounds of the distance from the victim to the adversary. As a result, the region in which the victim is locating is easily obtained by employing the trilateration model again, but with two circles drawing from the adversary to the two nearest neighbors as shown in Figure 4.

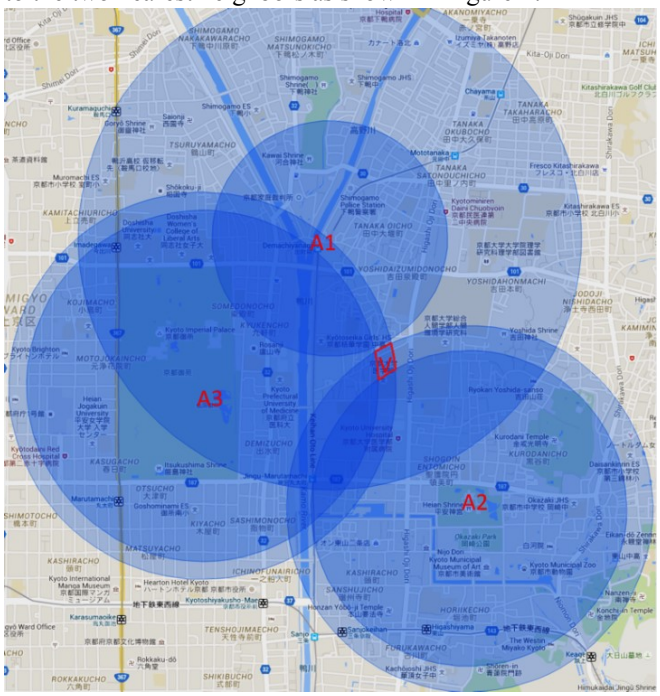


Fig. 4 Enhanced version of Trilateration Threat Model.

By using this model, the adversary even does not need to view the victim profile three times to record the distance as done in the original trilateration model, but still very effective at locating the victim's region as marked in Figure 5.

Fig. 5 The victim's region is smoothly bounded.

Hence, instead of querying the distance information of the victim several times, the adversary just needs to query it from the two nearest neighbors (*appear on the screen of Grindr*) of the victim from three different points, once for each. As a result, no distance query from the adversary to the victim is issued in this enhanced model. Therefore, approach like limiting the number of queries issued to get the distance information from one user to another user is not adequate in this case.

Moreover, a lesson learned from Figure 4 and Figure 5 is that the adversary gains more location information about the victim at A2 and A3 than A1, because the bounds set by the pairs of victim's neighbors from the viewpoints of A2 and A3 are narrower than from A1. Therefore, even without drawing the pair circles from A1, the adversary can still confidently infer the possible region of the victim, because one of the two intersection areas is a river space, thus the probability that the victim is locating in that region is relatively low. Since discussing about the probability of possible activity region of smart phone user is beyond the scope of this paper; for more information, the reader can refer to [13], in which the probability of possible locations of smart phone user has already been discussed.

So far, one may think that this type of attack model is not valid in case the victim locates in low-density area, because there are not so many neighbors around him for this attack model to take place. However, as far as we are concerned, using the term "victim's neighbors" actually is not always correct, because they may not physically locate near to the victim in real world. In fact, this attack model is still valid as long as the following condition holds:

$$AN1 < AV < AN2$$

Where:

- AV is distance from the adversary to the victim.
- AN1 and AN2 are distances from the adversary to the pair of the so-called victim's nearest neighbors.

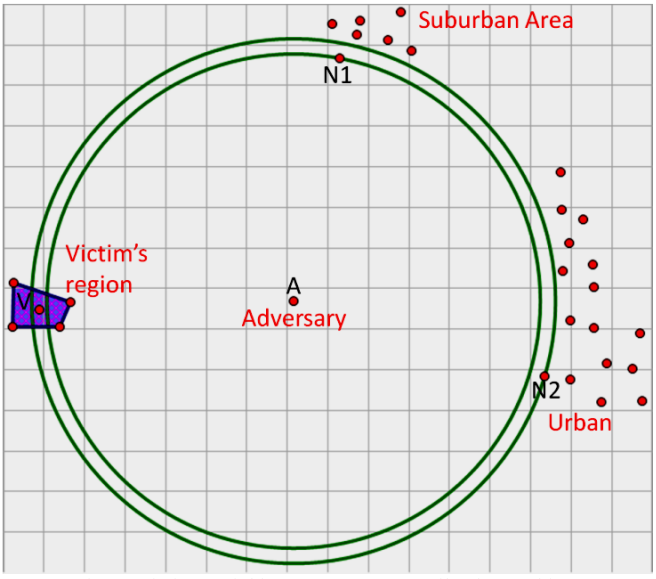


Fig. 6 Victim's neighbors are not necessarily close to him.

In real life attack, the adversary can apply this model to attack the victim in remote area by placing himself in the middle of high-density areas and the victim's region as shown in Figure 6. The more crowded the urban areas are, the more resources that the adversary can obtain to precisely explore the victim's location. Or, in a more active attack fashion, the adversary can create two colluding accounts and move them around until he can satisfactorily compromise the victim's location. The key idea is to gradually reduce the subtraction

value of |AN1-AN2| such that V is still sandwiched by N1 and N2 on the application's screen of the adversary. The smaller the subtraction value becomes, the more accurate location of the victim can be revealed. Up to this point, it is obvious that obtaining victim's real location from Figure 5 becomes a trivial task. With this enhanced model, even when all local members hide their distances, the adversary can still make completely use of his colluding fake accounts to infer, thus be able to narrow down the possible region in which the victim is locating. That is where the name "colluding trilateration" of our proposed method originates from. The idea is demonstrated in Figure 7.

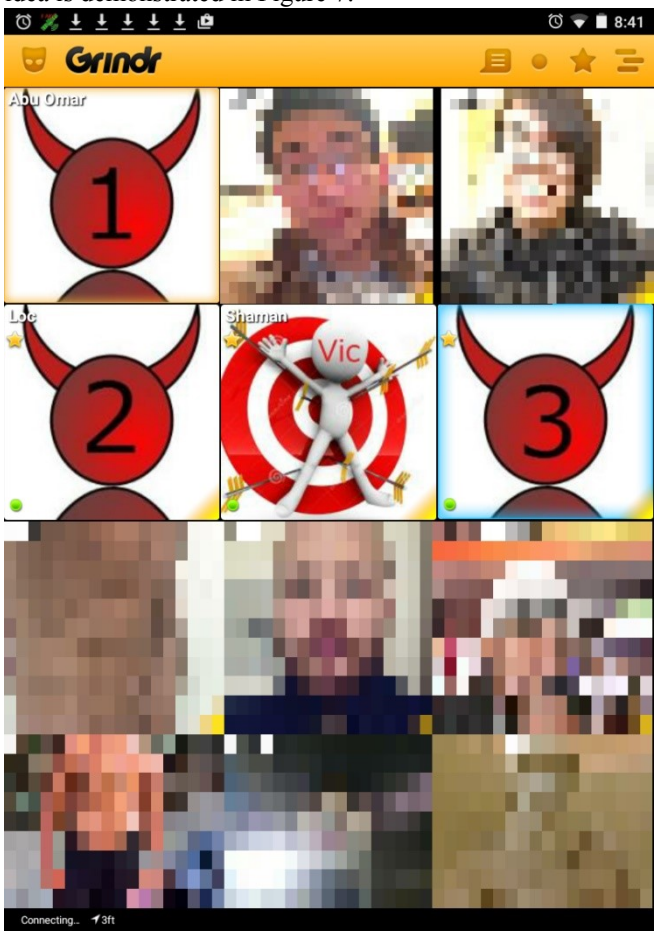


Fig. 7 Colluding Trilateration in Grindr.

As illustrated, the adversary, whose account is marked with 1, can create two colluding accounts 2 and 3, then positions 2 and 3 in a way that the victim (*Vic*) is sandwiched between 2 and 3. Later, the adversary can gradually move 2 and 3 so that the value of the distance between them becomes smaller while the victim remains in between of 2 and 3 on the screen, thus be able to precisely figure out the distance of the victim basing on the distances shown in the profile of the colluding accounts 2 and 3.

C. *Location Obfuscation is not enough*

While Grindr adopts location anonymization approach to protect its users from the trilateration threat model, Hornet adopts both location anonymization and location obfuscation. In other words, all of the publicly shown distances in Hornet are obfuscated as mentioned in our preliminary result above. Even from a same location, we keep receiving different distance values every time we issue a new query to reload the profile page of the victim. Therefore, we can initially confirm that Hornet does not use a one-to-one function to obscure the real distance. In addition to this feature, Horner user can also choose to hide his obfuscated distance from other users. As a result, Hornet seems to be better in protecting its user's location from the attack fashion of trilateration model.

Indeed, Hornet was first released in 2011[3], later than Jack'd (2010)[4] and Grindr (2009)[5], thus carefully designed with the concern of privacy and security issues [14].

Because Grindr is one of the first GLBT-focused applications introduced to the society, and widely used in North America, it has become a typical object for many studies ranging from privacy, psychology, gender to sexuality health and so on; while Hornet is new and has not been thoroughly studied, especially in the aspect of user privacy and information security. For instance, Hornet was recently discussed in [15], but the authors did not accurately study the distance obfuscation pattern of Hornet. Instead, they stated that Hornet sends the distance with 10 m accuracy. However, it is not true, since Hornet does carefully obfuscate the distance of its user as stated on its homepage [16]. Therefore, to the best of our knowledge, we are the first group that empirically studies the location privacy aspect of Hornet.

In order to discover the location obfuscation pattern of Hornet, we repeatedly move around all of our adversary accounts to about 3000 different locations within 3 Km from our institute, and record both real distances estimated by ourselves and obfuscated distances shown in our Hornet accounts. Next, we scatter the data and obtain the graph in Figure 8.

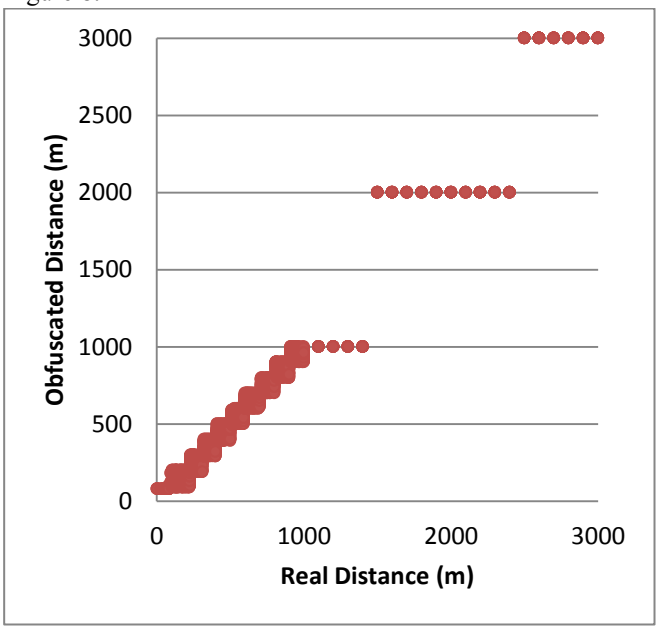


Fig. 8 Hornet Location Obfuscation Pattern.

As we already noticed that Hornet does not simply use a one-to-one function to obscure the distance, we cannot just easily employ the regression analysis to figure out the location obfuscation pattern used in Hornet. By observing the recorded data, we found that Hornet is very careful in designing its location obfuscation scheme with three different strategies for three different ranges of distance, which are 0~100 m, 100~1000 m, and above 1 km. Within the range of

[3] https://www.appannie.com/apps/all-stores/app/hornet/
[4] https://www.appannie.com/apps/all-stores/app/jackd/
[5] https://www.appannie.com/apps/all-stores/app/grindr/

0~100 m, Hornet makes the publicly shown distance equal to 80 m for all of the distances which are shorter than 80 m. For distance longer than 80 m and shorter than 100 m, Hornet arbitrarily adds some noise to the real distance such that the obscured distance varies within 80 m and 100 m. We break down the Figure 8 to have a clearer view on the distance obfuscation pattern in the range of 0~100 m.

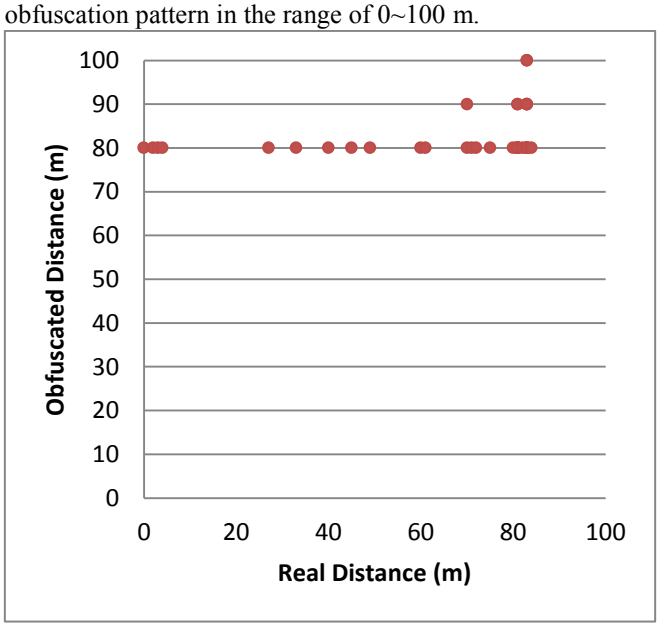


Fig. 9 Hornet Location Obfuscation Pattern – Range 0~100m.

Next, within the range of 100~1000 m, Hornet adopts a more complex obfuscation pattern that changes the publicly shown distance every time we refresh the victim profile to get the distance information. As a result, although the positions of our fake accounts did not change, we always got different values shown in our fake accounts. To deeply analyze this obfuscation pattern, we issue 30 queries from every single position and note down all the possible obfuscated results returned by Hornet. Breaking down the Figure 8 in the range of 100~1000 m gives us a clearer view.

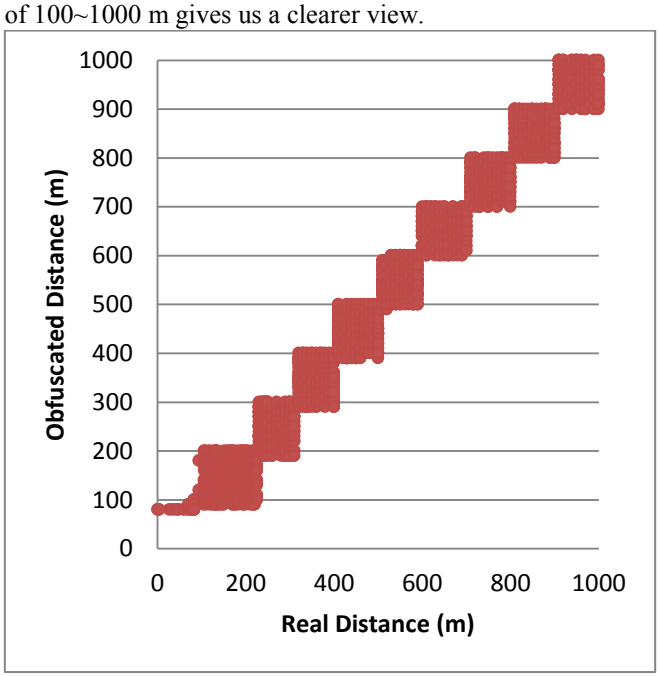


Fig. 10 Hornet Location Obfuscation Pattern – Range 100~1000 m.

From Figure 10, we can conclude that in the range of 100~1000 m, Hornet evenly randomizes the real distance with amplitude of 100 m. In more detail, a real distance, which is longer than 100 m and shorter than 1000 m, is first rounded to the nearest hundred. It is then obfuscated by adding an arbitrary number ranging from 0 to 100 with the step of 10. For example, if the real distance is 321 m, then the obfuscated distance is 300 plus any random value in the set of {0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100}. Thus, the publicly shown value can be any number between 300 m and 400 m with the step of 10.

Finally, for distance longer than 1 km, Hornet applies another obfuscation pattern in which the real distance is rounded to the nearest one in the unit of km. For instance, 1.2 km is rounded to 1 km, while 1.6 km is rounded to 2 km.

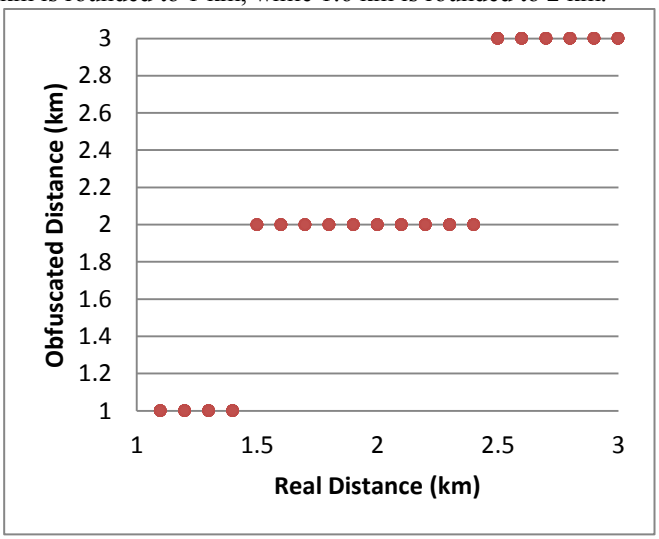


Fig. 11 Hornet Location Obfuscation Pattern – Range 1 km ~.

From the patterns discussed above, it is not an exaggeration to say that the obfuscation used in Hornet is very reasonable with respect to the privacy preserving for Hornet user. Distance of every user in the range of 0~1000 m is complexly obfuscated because this range is considered to be very sensitive to the user's location privacy; while distance longer than 1 km is simply obscured by rounding method because it is far enough to protect the user privacy. Also, showing the approximated distance for those users, whose distance is longer 1 km, does not have any bad impact on the utility of other local users. Therefore, the obfuscation does not only provide Hornet user with a better protection, but also makes it more difficult for the adversary to carry out the trilateration attack. That is the reason why we could not find out the exact location of the victim in Hornet when applying the trilateration model because three circles sketched from three different locations of our adversary accounts do not converge on one point.

Nevertheless, we still have the colluding-trilateration method proposed in Subsection II.B to test in Hornet; because, same as Grindr, Hornet also arranges the users on the screen from left-to-right and top-to-down in an ascending order of their distances regardless of whether they have disabled the "show distance" function, and the publicly shown distance has already been obfuscated. Therefore, we reuse the colluding-trilateration to examine whether we can discover the real location of the victim or not. This time, we just move accounts 2 and 3 to obtain a same result as shown in Figure 7, but do not use the distances shown on these two fake accounts because they are already obfuscated by Hornet, thus not correct. However, because those fake accounts

belong to us, we can always measure the real distances between them as side-channel knowledge without relying on the one shown in Hornet.

Surprisingly, Hornet perhaps might have already envisioned about this type of attack and took a step ahead. What Hornet does is to randomly remove some users from the screen of other local users. In other words, at some positions, on the screen of account 1, we can see all of the victim (*Vic*) and accounts 2 and 3; but at some other positions we cannot. That means Hornet does not always show all surrounding local users on one's screen. Instead, it arbitrarily drops some users. As a consequence, we cannot set the upper and lower bounds for the victim's distance to conduct the colluding-trilateration.

In order to bypass this countermeasure, we make completely use of the Favorites List of Hornet. Similar to most of other Social Network platforms, Hornet also has a feature in which a user can add other users to his favorite list like the "follow" function in Twitter. Then, there is a separate tab for the user to view his Favorites List in which the favorite users in this list are also arranged from left-to-right and top-to-down in an ascending order of their distances regardless of whether they have disabled the "show distance" function, the publicly shown distance are obfuscated, or they are not shown on the local screen of other users. The trick is demonstrated in Figure 12. (*Two accounts at the top-left corner are added by default for all users in Hornet to promote sexuality health.*)

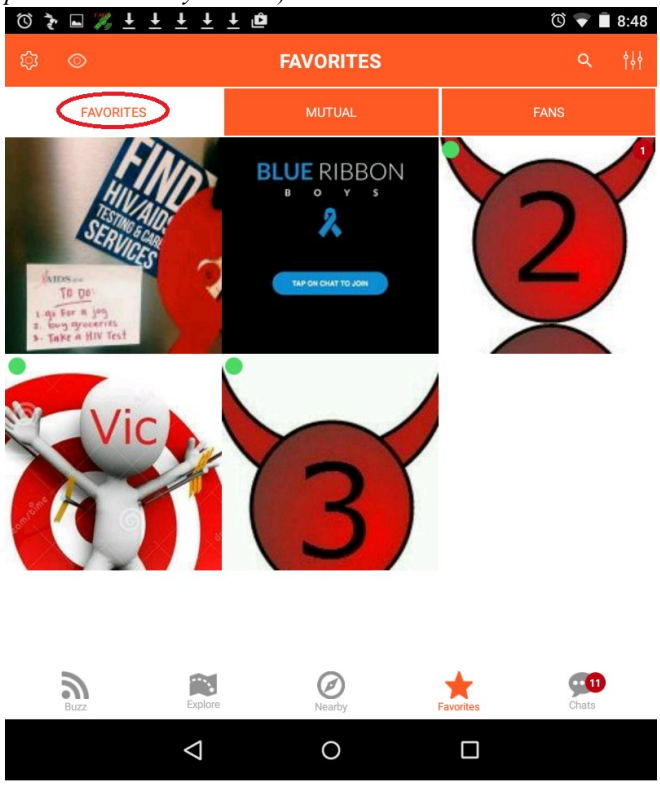


Fig. 12 Colluding Trilateration in Hornet bases on Favorites List.

Taking advantage of this Favorites List, the adversary first needs to add the other two fake accounts 2 and 3 to his Favorites List, and then also adds the targeted victim into this list at the first time he sees the targeted victim on the public local screen. That is how the adversary can anchor the victim even when the victim does not appear on his local screen. Finally, the locating problem becomes a same problem as shown in Figure 4. As a result, we were able to locate the position of the victim (*Vic*) in Hornet although all the distances are obfuscated and the victim already disabled the "show distance" function in his application.

III. OTHER PRIVACY CONCERNS

A. Vulnerabilities from Third-Party Advertisement Banner and Misconduct in Handling User's Personal Information

In the age of online marketing, one's privacy is often threatened by the very advertisements popping up in his device as mentioned in [17] and [18]. In order to investigate this issue in Jack'd and Grindr, we analyze packets captured while using these two applications. The experiment results are shown in Figure 13.

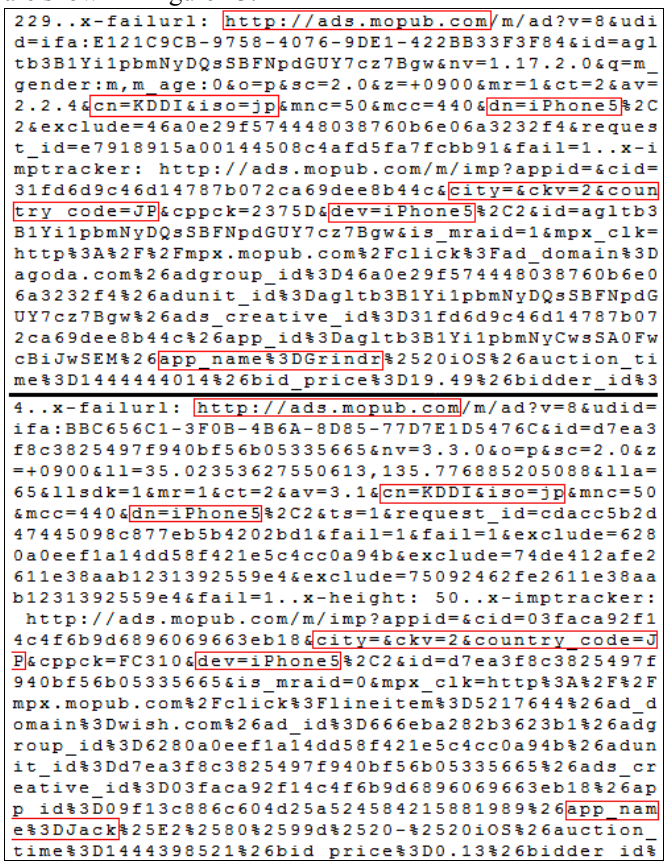


Fig. 13 : Information Leak through Third-Party Ads.

Surprisingly, in both applications, the third-party advertisements leak many important information of the user including name of Telecommunications Service Provider (i.e. KDDI), device's model information (i.e. iPhone 5), country code (i.e. JP), and last but not least: the name of the applications (i.e. Grindr and Jack'd) which is the most sensitive information that no any straight-acting person wants the others to know the existence of such applications in his phone. It may have no problem if the packets are sent directly to the ads provider's server. However, what is worth mentioning here is that the packets are sent in an unencrypted fashion, thus widely open to an attack type known as man-in-the-middle attack, in which the hacker taps the Internet connection of the victim to eavesdrop the packets.

By analyzing all the captured packets, we were further shocked by the fact that all the packets containing members' profile pictures of Grindr are also sent in the air without encryption, thus being captured and recovered back to the original image files as shown in Figure 14.

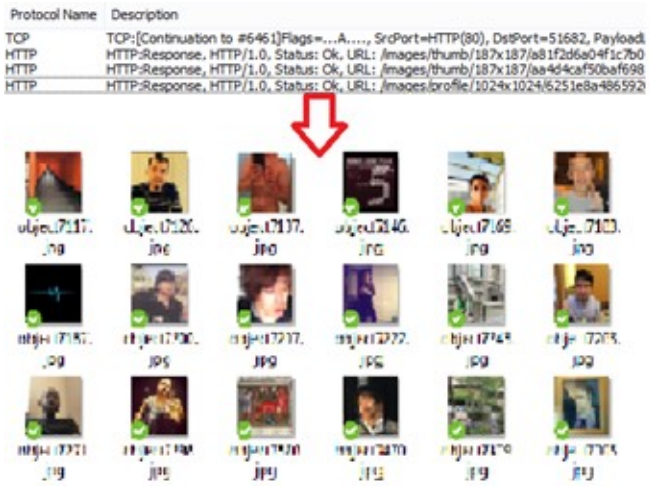


Fig. 14 Pictures recovered from unencrypted packets in Grindr.

Concerning ethical issue, we only recover image files that appear on the first screen page of Grindr as evidence, and the users' avatars are intentionally censored. For Jack'd, it is very careless in handling its user's private photo, because even when a photo is sent in a private message, Jack'd does not use any secure connection to protect the photo. Instead, Jack'd sends it via HTTP, which is an unsecure transmission protocol as shown in Figure 15.

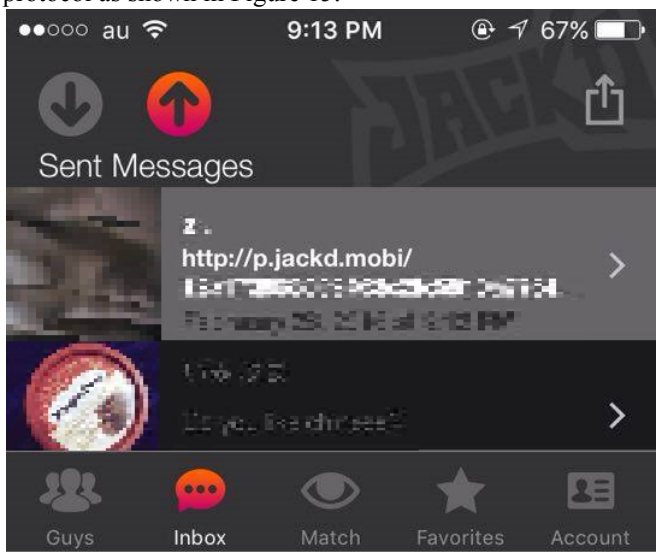


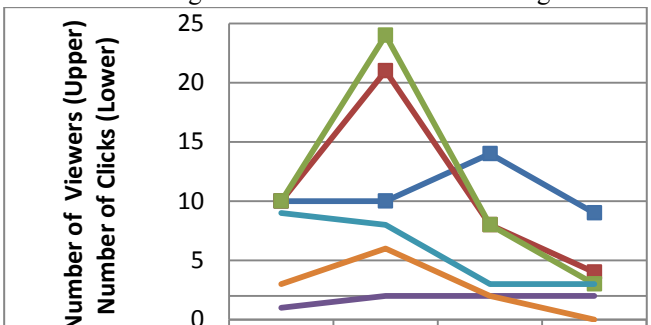
Fig. 15 Unsecure http protocol used in Jack'd to send private photos.

Taking into account of the above findings, it is obvious that the user's privacy is not guaranteed at all although the vendor has been alerted to these issues by a security firm before [11].

As for Hornet, it is not necessary to do this experiment, because it has been already confirmed in [14] and [15] that Hornet carefully employs SSL certificates and HTTPS protocol for its connection.

B. *Together with IP Spy and Linkage Attack*

Next, as human being is born curious, social engineering intrusion techniques like phishing is always the easiest but effective way to compromise people. Since Jack'd provides it users with a feature to see who viewed his profile with timestamp, an adversary can put an IP-spy URL into his profile to promote his appearance, thus being able to obtain the victim's IP address if the victim feels curious and clicks on that URL. Nevertheless, as it was discussed in [19] that IP address is also important personal information which can be exploited to perform the linkage attack to retrieve other personal information. To have a clearer view on how this gimmick is really effective at luring innocent users, we place our Jack'd accounts in three big cities of Japan which are Tokyo, Osaka and Kyoto within 12 hours (*from 6PM to 6AM of the following day*) to estimate how many innocent and curious victims could be lured. We choose this time period because it is the most active usage time according to [20]. To conduct this task, we had to reboot the virtual machines every two hours so that our accounts will not disappear from the screen of other local users, as we found that Jack'd only keeps an account displayed on other user's screen for two hours since the latest login. The result is illustrated in Figure 16.



	18～21	21～00	00～03	03～06
Tokyo	10	10	14	9
Osaka	10	21	8	4
Kyoto	10	24	8	3
Tokyo-click	1	2	2	2
Osaka-click	9	8	3	3
Kyoto-click	3	6	2	0

Fig. 16 Analysis from IP-spy Intrusion Gimmick.

In total, we got 131 viewers from three accounts with 41 viewers clicked through the IP-spy URL we put in the profile.

Fig. 17 Linkage Attack with IP-spy and Jack'd timestamp.

Among these 41 clicks, we were able to perform linkage attack to 26 users with high confidence by matching the timestamp between Jack'd and our IP-spy server to further reveal other information including their IP address, ISP, display language and platform of their devices as shown in Figure 17.

IV. DISCUSSION AND CONCLUSION

Through this study, we would like to particularly alert the users of Grindr, Jack'd, and Hornet as well as the users of other LBS in general about the risk of being located easily regardless of whether the recent location anonymization and location obfuscation approaches have been adopted. By investigating these three applications, we found a paradox that although there have been many attack models proposed by the privacy-preserving researchers, the user's location privacy has not been seriously taken in to consideration by the LBS provider and the user themselves. As far as we are concerned, the reason of this negligence derives from both sides. From the viewpoint of the LBS provider, it might cause overhead to implement those sophisticated solutions proposed by the research community, while the utility of the application is not really guaranteed, thus probably lead to the loss of its customer. From the viewpoint of the user, it is maybe because of two reasons. Perhaps, the first one is also due to the trade-off of utility. The other one is because of unawareness. For that reason, in this paper, instead of using complicated mathematical equations and complex algorithms to show the threat models, we opt for visualizing it on maps and figures so that even those non-technical readers can understand how easily their privacy can be compromised in the current security condition.

In order to alleviate the risks of man-in-the-middle attack, IP spy and other side channel attacks as mentioned in Section III, we urge the LBS providers and involved third-parties to carefully encrypt the connection from their servers to the users. For the user, we suggest not opening any URL out of curiosity. If it is really necessary to open an unknown URL, the user should turn on VPN at first to prevent the leak of their real IP address.

For those threat models discussed in Section II, let us argue that privacy preserving policy is different from person to person. Especially in GLBT community, some already came out, thus have no concern about privacy; while some are straight-acting, thus do not want to be disclosed. Therefore, a centralized solution is not really suitable, and users are the very ones who need to make decision whether to protect their own privacy. For the meantime, while waiting for the experts and the LBS providers to discover a perfect solution for location privacy protection without trading off the utility, we suggest that the user should take a step ahead to protect their own privacy from those vulnerabilities mentioned in this study. That is to use Fake-GPS applications like the one that we use in this study (*probably also used by most of the adversaries*) to hide the real location to an acceptable extent so that the user can still gain the convenience provided by the LBS. How far the fake location should be shifted from the real one depends on how much utility and convenience that a user is willing to trade off with his privacy, thus different from case by case. We strongly believe that this user-centric solution not only suits all type of users, but also helps to save the vendors from overhead investment in implementing sophisticated solutions and infrastructures.

Apart from technical methods aforementioned, human factor is also important in protecting oneself in the cyberspace. In order to avoid troublesome problems in the future when the vendors get hacked as the case of Ashley Madison [6], the user should not register account to those highly sensitive applications under his real name or even a part of his real name. Instead, the user should use information that could not be used to link the account with his real-life personally identifiable information.

Last but not least, in this paper, we of course could have utilized more complicated techniques to extract and test the accuracy of the threat models with more users of Grindr, Jack'd, and Hornet in bulk. However, as far as we are concerned with the ethical issue that those compromised users also have their right to be undisclosed, and there may be our acquaintances among them. We thus did not go beyond those accounts created by ourselves.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number 15K00423 and the Kayamori Foundation of Informational Science Advancement.

REFERENCES

[1] P. Jose, "Tim Cook: Privacy is a matter of 'life and death,'" *CNN*, 2015. [Online]. Available: http://money.cnn.com/2015/02/13/technology/security/tim-cook-apple-cybersecurity/. [Accessed: 20-Sep-2015].

[2] Ponemon Institute, "2015 Cost of Data Breach Study: Global Analysis," IBM Corporation, 2015.

[3] B. Schilit, J. Hong, and M. Gruteser, "Wireless Location Privacy Protection," *Computer (Long. Beach. Calif).*, vol. 36, no. 12, pp. 135–137, Dec. 2003.

[4] M. Li, H. Zhu, Z. Gao, S. Chen, K. Ren, L. Yu, and S. Hu, "All your location are belong to us: Breaking mobile social networks for automated user location tracking," in *MobiHoc*, 2014, pp. 43–52.

[5] Y. Ding, S. Peddinti, and K. Ross, "Stalking Beijing from Timbuktu: A Generic Measurement Approach for Exploiting Location-Based Social Discovery," in *Proceedings of the 4th ACM Workshop on Security and Privacy in Smartphones & Mobile Devices*, 2014, pp. 75–80.

[6] Staff CNNMoney, "The Ashley Madison hack...in 2 minutes," *CNN*, 2015. [Online]. Available: http://money.cnn.com/2015/08/24/technology/ashley-madison-hack-in-2-minutes/. [Accessed: 20-Sep-2015].

[7] ILGA, "State-Sponsored Homophobia," 2014.

[8] H. DEBORAH, "ISIS terrorists pose as gay men, lure victims on dates, then kill them," *NEW YORK DAILY NEWS*, 2015. [Online]. Available: http://www.nydailynews.com/news/world/isis-terrorists-lure-gay-men-deaths-article-1.2197555. [Accessed: 20-Sep-2015].

[9] E. Mishel, "Discrimination against Queer Women in the U.S. Workforce: A Resume Audit Study," *Socius Sociol. Res. a Dyn. World*, vol. 2, Jan. 2016.

[10] T. D. Cook, "Tim Cook Speaks Up," *Bloomberg*, 2014. [Online]. Available: http://www.bloomberg.com/news/articles/2014-10-30/tim-cook-speaks-up. [Accessed: 15-Feb-2016].

[11] "Grindr: A chronicle of negligence and irresponsibility.," *Synack, Inc.*, 2014. [Online]. Available: https://grindrmap.neocities.org/. [Accessed: 20-Sep-2015].

[12] "Grindr Security," *Grindr Team*, 2014. [Online]. Available: http://www.grindr.com/blog/grindr-security. [Accessed: 20-Sep-2015].

[13] Y. Michalevsky, A. Schulman, G. A. Veerapandian, D. Boneh, and G. Nakibly, "PowerSpy: Location Tracking Using Mobile Device Power Analysis," in *24th USENIX Security Symposium (USENIX Security 15)*, 2015, pp. 785–800.

[14] "With Hornet, Gays Can Now Play Safe on Gay Mobile Social Networks," *PRNewswire*, 2012. [Online]. Available: http://www.prnewswire.com/news-releases/with-hornet-gays-can-now-play-safe-on-gay-mobile-social-networks-137800183.html. [Accessed: 15-Feb-2016].

[15] P. Constantinos, Z. Athanasios, and S. Agusti, "Analysis of Privacy and Security Exposure in Mobile Dating Applications," in *The First International Conference on Mobile, Secure and Programmable Networking (MSPN'2015)*, 2015, pp. 151–162.

[16] "Keeping Our Users Safe — Hornet Networks," *Hornet*, 2014. [Online]. Available: http://love.hornetapp.com/blog/2014/9/23/keeping-our-users-safe. [Accessed: 15-Feb-2016].

[17] Angelia and D. Pishva, "Online advertising and its security and privacy concerns," in *The 15th IEEE International Conference on Advanced Communication Technology (ICACT)*, 2013, pp. 372–377.

[18] T. Chen, I. Ullah, M. A. Kaafar, and R. Boreli, "Information leakage through mobile analytics services," in *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications - HotMobile '14*, 2014, pp. 1–6.

[19] N. P. Hoang and D. Pishva, "Anonymous communication and its importance in social networking," in *The 16th IEEE International Conference on Advanced Communication Technology (ICACT)*, 2014, pp. 34–39.

[20] W. C. Goedel and D. T. Duncan, "Geosocial-Networking App Usage Patterns of Gay, Bisexual, and Other Men Who Have Sex With Men: Survey Among Users of Grindr, A Mobile Dating App," *JMIR Public Heal. Surveill.*, vol. 1, no. 1, p. e4, May 2015.

[21] N. P. Hoang, Y. Asano, and M. Yoshikawa, "Your Neighbors Are My Spies : Location and other Privacy Concerns in Dating Apps Your Neighbors Are My Spies : Location and other Privacy Concerns in Dating Apps," in *the 18th IEEE International Conference on Advanced Communication Technology (ICACT)*, 2016, pp. 719–725.

[22] "Top Five Most Gay Apps For iPhone and Android," *Clapway*, 2016. [Online]. Available: http://clapway.com/2016/01/12/top-five-most-gay-apps-for-iphone-and-android/. [Accessed: 15-Feb-2016].

Nguyen Phong HOANG was born in Tien Giang Province, Vietnam in 1992. He received his undergraduate degree in Business Administration majoring in Information & Communications technology (ICT) from Ritsumeikan Asia Pacific University, Japan. He is presently pursuing his graduate studies at the Graduate School of Informatics, Kyoto University, Japan. His research interests include information security, privacy and anonymous communication. He hopes to advance his research on Tor (The Onion Router), one of the most robust anonymous tools, during his graduate studies. He has participated in annual IEEE International Conference on Advanced Communication Technology (ICACT) since 2014. He also received the Outstanding Paper Award from the Technical Program Committee of the conference in the 16th and 18th ICACT. He has been an IEEE member since 2013, and DBSJ since 2014.

Yasuhito ASANO received the BS, MS, and DS degrees in information science from the University of Tokyo in 1998, 2000, and 2003 respectively. In 2003-2005, he was a research associate in the Graduate School of Information Sciences, Tohoku University. In 2006-2007, he was an assistant professor in the Department of Information Sciences, Tokyo Denki University. He joined Kyoto University in 2008. He currently serves as an associate professor in the Graduate School of Informatics. His research interests include web mining, network algorithms. He is a member of the IEICE, IPSJ, DBSJ, and OR Soc. Japan.

Masatoshi YOSHIKAWA received the BE, ME, and PhD degrees from the Department of Information Science, Kyoto University in 1980, 1982, and 1985, respectively. From 1985 to 1993, he was with Kyoto Sangyo University. In 1993, he joined the Nara Institute of Science and Technology as an associate professor in the Graduate School of Information Science. From April 1996 to January 1997, he was in the Department of Computer Science, University of Waterloo as a visiting associate professor. From June 2002 to March 2006, he served as a professor at Nagoya University. From April 2006, he has been a professor at Kyoto University. His current research interests encompass database technologies and their application to medical healthcare domains. He is a member of the ACM, IEICE, IPSJ and DBSJ.

Volume 5 Issue 3, May. 2016, ISSN: 2288-0003

ICACT-TACT JOURNAL

Global IT Research Institute

1713 Obelisk, 216 Seohyunno, Bundang-gu, Sungnam Kyunggi-do, Republic of Korea 13591

Business Licence Number : 220-82-07506, Contact: secretariat@icact.org Tel: +82-70-4146-4991