A Comparative Study of Privacy Protection Methods for Smart Home Environments

Homin Park, Taejoon Park and Sang Hyuk Son

Department of Information and Communication Engineering Daegu Gyeongbuk Institute of Science and Technology, Daegu, Korea {andrewpark, tjpark, son}@dgist.ac.kr

Abstract

Due to the importance of privacy protection for smart home environments, there have been numerous propositions, to support event source anonymity using fake data transmissions and delays. Among them, periodic and probabilistic distribution-based transmission methods achieve near perfect privacy while sacrificing real time transmission of detected event data. In this paper, we aim to study the performance of various privacy protection methods for smart home environments. Our result demonstrates that 1) periodic transmission method achieves perfect privacy while resulting in large average latency of real data transmission, 2) probabilistic distribution-based transmission method lessen the latency issues of periodic method while guaranteeing near perfect privacy, and 3) privacy level depends on the number of fake data transmission when real data is triggered immediately. Based on the evaluation result, we conclude that there is trade-off between privacy, latency, and energy efficiency.

Keywords: privacy for smart home environments, latency, energy efficiency, trade-offs

1. Introduction

Smart home environments, based on wireless sensor networks (WSN), can benefit from the key convergence technologies to provide autonomous assistance for various needs of the residents. These smart home environments are typically utilized by assisted living facilities where the residents receive specific medical care through activity monitoring. However, despite their benefits and broad applicability, there are critical privacy challenges arising from the characteristic of WSN that allow intelligent adversaries to eavesdrop and accumulate wireless data transmissions to infer the residents' Activity of Daily Living (ADL). This attack scenario is serious because ADLs may convey very private and personal information of the residents [7]. One of the most dangerous privacy attacks based on statistical inference is the Fingerprint And Timing-based Snooping (FATS) attack [1]. It was successfully demonstrated that using FATS attack, ADLs of the residents could be inferred regardless of message encryption techniques and diversity of facility layouts.

To protect the privacy of the residents in smart home environments, there have been extensive studies on event unobservability and source anonymity¹ concept, which obscures real ADLs via fake data transmissions with intentional delays on real data transmissions [2, 3, 4, 6]. Clearly the highest level of privacy is attained by periodically transmitting messages, which guarantees perfect privacy preservation since every sensor node in the network will

Taejoon Park is the corresponding author.

¹ Unobservability is the state of items of interest (IOI) being indistinguishable from any IOI at all, while anonymity is the state of being not identifiable within a set of subjects, called the anonymity set [10].

have the same transmission pattern, making it very difficult for the adversaries to distinguish individual sensor from the others. On the other hand, the periodic transmission method suffers large amounts of latency since detected events must be buffered until the next scheduled transmission time. Such latency of real event reporting causes degradation on quality of service (QoS) in many applications, such as cardiac activity and blood pressure monitoring system used in smart home environments. It also incurs additional data traffic increasing the overall energy consumption since the data has to be transmitted at every scheduled time regardless of whether or not an event was detected, leading to the transmission of meaningless data if there is no real data in the buffer.

To resolve the latency issue of periodic transmission method, Shao et al. introduced a probabilistic distribution-based transmission (PDT) method [2]. The key idea of this method is to schedule transmission intervals that are random in length based on a probabilistic distribution, ensuring complete randomness of the entire traffic. Compared with periodical transmission, PDT minimizes the latency of real data transmissions through much frequent transmission schedules, while ensuring near perfect privacy. On the other hand, although such protection method achieves minimum latency, utilizing frequent fake data transmissions causes energy efficiency issues. The problem of energy efficiency is critical because most sensor nodes are battery-powered and the energy cost for data transmission is one of the most significant ones. In fact, frequent fake data transmissions are not favorable for the longevity of the network. On the other hand, if we consider immediate transmission of real data for optimal QoS for smart home environment applications, the privacy of the resident would be totally compromised by the adversaries. To protect privacy of the residents while real data transmissions are made with no delays, we can adopt source simulation concept introduce by Mehta [8] where fake data transmissions are made following the PDT method to obscure real data transmission patterns.

In this paper, we aim to study the performance of three privacy protection methods for smart home environments, namely, periodic, PDT, and immediate transmission methods. We evaluate and compare the performance of these methods using the following performance criteria as a function of the ratio of fake to real data transmissions: 1) privacy threat level, 2) latency, 3) energy efficiency, and 4) Anderson-Darling (AD) test acceptance percentage. In order to simulate aforementioned methods comprehensively, we utilize both handcrafted and real data set obtained from [9].

The rest of this paper is organized as follows. Section 2 discusses the background information of smart home environment privacy. Section 3 presents the assumptions that we have made as well as the discussions on the evaluation criteria. Section 4 provides comparison results of privacy protection methods in smart home environments, and Section 5 concludes this paper.

2. Background

2.1. Privacy Threat in Smart Home Environments

Because smart home environments inherit vulnerabilities of WSN such as packet eavesdropping, most of the traditional protective measures cannot fully address the privacy issues [1]. For example, if intelligent adversaries eavesdrop and accumulate timestamps of data transmissions over the wireless network, they can easily identify when residents are occupying the facility. In addition, because each sensor node has a unique radio wave pattern (called transmission fingerprint), analyzing transmission patterns according to transmission fingerprints can reveal very sophisticated ADL information about the residents. The FATS attack introduced by Srinivasan, *et al.*, [1] is one of the most dangerous privacy inference

attacks targeting smart home environments. By analyzing accumulated fingerprints and timestamps of transmitted data, the adversaries are capable of identifying behavioral patterns of the residents regardless of the facility layouts and encryption techniques. As shown in Figure 1, FATS attack utilizes a multi-tier algorithm to identify and classify sensor allocations as well as their functionalities, which leads to the inference of ADLs of residents. The algorithm consists of 4 tiers that are designed to achieve specific objectives as described below.

Tier-0. detects if residents are at home, away or sleeping.

- **Tier-1.** performs sensor clustering to identify the number of rooms and a list of sensors belonging to each room.
- Tier-2. classifies the rooms to identify the room types, such as bathroom or kitchen.
- **Tier-3.** classifies the sensors within each room to precisely capture their activities like showering or cooking

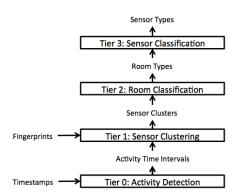


Figure 1. Multi-tier Algorithm for FATS Attack

As shown in Figure 1, each tier of FATS attack yields specific results that are utilized by the next tier to infer more sophisticated and fine-grained information. In tier 1, the intelligent adversaries utilize fingerprints of each sensor node to identify unique data transmission patterns. These transmission patterns are then used to identify which sensors are allocated in each room, under the assumption that sensors in the same room typically generate similar transmission patterns because they together detect the same activities performed in the room. In fact, tier 1 calculates the temporal distance of sensor nodes in the network in order to cluster sensors that are temporally close together. The clustering algorithm is shown in Figure 2, where ID is the set of all sensor nodes, and the vector of all transmission timestamps from each sensor node i = ID to be T_i . The algorithm also uses $T_i[k]$ to refer k^{th} timestamp from sensor node i. In Figure 2, SPDIST indicates Dijkstra's shortest path algorithm, CMDS refers to classical non-parametric multidimensional scaling, and k-means is associated with kmeans clustering algorithm. Details of the clustering algorithm can be found in [1].

The result of sensor clustering are then used by tiers 2 and 3 with sophisticated classification means to identify the room and sensor types, which can be used to infer ADLs of the residents. For intelligent adversaries, accuracy of the sensor clustering results become very important for attaining precise ADL information since classification results are heavily dependent on how sensors are clustered together. The evaluation results indicate that FATS

attack is able to identify ADLs of the resident with more than 90% accuracy in various environment settings [1].

Figure 2. Clustering Algorithm from Tier 1 of FATS Attack

2.2. Privacy Protection in Smart Home Environments

The privacy concerns in smart home environments stems from the fact that sensors are transmitting data without paying much attention to the possibility of disclosing actual ADL patterns of the residents. To protect the privacy of the residents, there have been extensive studies on providing event unobservability and source anonymity, which rely on superimposing fake data packets as well as delaying of real data packets. These studies demonstrated that the highest level of privacy could be achieved by periodically transmitting the data, which guarantees perfect privacy preservation since every sensor node in the network will have the same transmission pattern. On the other hand, delay of a real data transmission is inevitable if an event was detected before the next scheduled transmission time. Such latency of real data transmission is one of the critical factors that degrade QoS of smart home environment services.

To resolve such drawback of periodic transmission methods, Shao et al. proposed the PDT method [2]. This method schedules the transmission intervals that are random in length according to a certain probabilistic distribution. When event was detected, the real transmission is delayed until the time interval is fitted to predetermined probabilistic distribution. Consequently, PDT method achieves small transmission latency compared with periodic method while providing near perfect privacy. The degradation of privacy level is due to the fact that intelligent adversaries who are capable of analyzing random traffics over the entire network might have a chance to identify slight distinction between each sensor node traffic pattern using statistical inference algorithms, which analyzes transmission frequency and the time intervals. Nevertheless, attaining any meaningful information by analyzing the random traffic pattern is very costly, which result in providing desirable degree of privacy protection. On the other hand, although such protection method achieves small latency, utilizing frequent fake data transmissions significantly increases the energy consumption. The problem of energy efficiency issue is due to the fact the sensor nodes have very limited energy resources while radio transmission cost is most significant among other processes. In fact, frequent fake data transmissions reduce the lifetime of the network in direct proportion [4, 5]. By contrast, minimizing the fake data transmissions for energy efficiency without thorough consideration on latency will degrade QoS significantly.

With different perspective on privacy protection from previously discussed methods, Mehta et al. proposed a source simulation method for WSN [8]. This protection method assumes that the intelligent adversaries are capable of eavesdropping entire network with abundant recourses and interests. The main idea behind source simulation is to generate highly accurate behavioral model of interested entity, such as residents for smart home environments, which can mimic the actual behaviors of real person. By employing generated model, we can deceive the adversaries to believe there is more than actual number of residents in the facility, which leads to obscurity of real ADLs [8]. The downside of such method has to do with difficulty of generating accurate behavioral model. Employing inaccurate behavioral model for the network is most likely to fail to provide desired level of privacy, since adversaries will be able to neglect such meaningless transmission patterns. On the other hand, the advantage of such method is that 1) real data transmission can be made without any delays, and 2) does not require entire sensor nodes to transmit fake data. For the immediate transmission method, we decide to obscure real data using PDT method while real data transmissions are made without delays. The performance of discussed privacy protection method was evaluated by analyzing the tradeoffs between privacy, communication cost, and latency.

3. Comparative Study of Privacy Protection Methods

Smart home environments can be designed with various network specifications. In this paper, we assume that smart home environments are built on top of a homogeneous network where all sensors have similar resources, such as computing power and energy capacity. Such network configuration is very common for various applications that are used in practice as well as in academia. In addition, we assume that smart home environments are consisted of single-hop network, where data transmissions from each node are delivered to the sink without passing through the neighbor nodes. We believe such assumption is reasonable since current wireless communication technologies used for sensor nodes can cover the entire area of the facilities. We also assume that the intelligent adversaries are capable of performing global eavesdropping over entire network in order to attain timestamps and fingerprints of real data transmissions. These adversaries are equipped with sophisticated privacy attack means, such as statistical inference and data mining techniques. Finally, we assume that adversaries are utilizing FATS attack to infer ADLs of the residents.

Under these assumptions, we have studied the effect of various privacy protection methods in smart home environments with following performance metrics: 1) privacy threat level, 2) latency, 3) energy efficiency, and 4) AD test acceptance percentage. These criteria will be discussed in the following sections.

3.1. Privacy Threat Level

For the purpose of measuring the privacy level that is provided by various protection methods against FATS attack, we have utilized the accuracy of clustering result coming from tier 1 of FATS attack to denote *privacy threat level*. The computation for clustering accuracy uses a maximal min cost bipartite mapping method between sensor clustering result and actual allocation of sensors and rooms based on the initial deployment plan. Based on such computation method, the accuracy of sensor clustering result is the proportion of sensor nodes that are correctly clustered. Referring back to Figure 2, the accuracy of sensor clustering result could be maximized when value k from k-means clustering algorithm matches with actual number of room in the smart home environment [1]. The privacy threat level has the following physical meaning. If the clustering result perfectly matches with actual allocation

of sensors and rooms, we denote privacy threat level is equal to 1 (the lowest level of privacy preservation) because it will enable the adversaries to completely compromise the privacy of the residents. In contrast, if the clustering result yields no meaningful information related to the actual sensor allocation, set the privacy threat level to 0 (the highest level of privacy preservation) because privacy is perfectly preserved. Simply put, the privacy threat level is directly proportional to the clustering accuracy, *i.e.*, the less relevant the clustering result, the safer the privacy of the resident, and vice versa.

3.2. Latency and QoS

Smart home environments are built to provide autonomous services to various needs of the residents. In fact, there exist various applications with different constraints on latency of real data transmissions. Some applications, such as cardiac activity and blood pressure monitoring system, need to satisfy real time constraints, because they cannot afford latency since QoS of these applications depends on how quickly the assists are given when problems are detected. For the privacy protection methods discussed in the previous sections, there exists a trade-off between latency and energy efficiency. In case of periodic transmission method, to reduce the latency of real data transmissions, we must decrease the length of the periodical time interval. Consequently, more and more transmissions are required which directly affects the energy efficiency of the network. In contrast, if we increase the length of the time interval for energy efficiency, longer delays of real data transmission are inevitable. To resolve the latency issues of periodic methods, PDT method was introduced and achieved very small latency using random time intervals based on a probabilistic distribution. On the other hand, considering the work that was done by Mehta, et al., [8], we have observed that privacy of an entity can be also protected even when the real data transmissions are immediately triggered when event is detected by using source simulation model.

3.3. Energy Efficiency

The energy efficiency issues in smart home environments are critical because energy resources are limited while energy cost in radio transmission is significant. In order to enhance the energy efficiency, minimizing the number of data transmission is very important. For the purpose of evaluating various privacy protection methods, we have quantified the energy efficiency by utilizing a ratio of fake to real data transmissions. Intuitively, as the fake-to-real ratio increases due to the increase in fake data transmission number, the energy efficiency will get worse because of the extra energy consumed by transmitting meaningless information, and vice versa.

3.4. Anderson-Darling Test Acceptance Percentage

The AD test is a statistical test that evaluate whether the time intervals used for scheduling data transmissions are drawn from a same probabilistic distribution. Considering the immediate real data transmission that was discussed in the previous section, real data transmission patterns are not likely to follow a specific probabilistic distribution since behaviors of the residents are independent from mathematical means. Such phenomenon is very dangerous for residents' privacy since intelligent adversaries can distinguish the real data transmissions from the fake data transmissions when the time interval fails to pass the AD test. In case of periodic and PDT method, AD test becomes meaningless since time intervals from these methods strictly follow a certain distribution. Therefore, AD test acceptance percentage is used to evaluate the feasibility of immediate transmission method where real data transmissions are made without delays.

4. Performance Evaluation

4.1. Simulation Environment

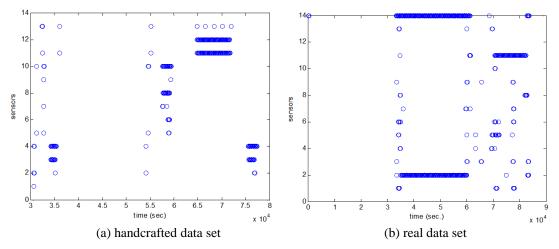
To evaluate the privacy protection methods under consideration, we have utilized two different data sets, one is handcrafted, and the other is actual transmission data obtained from [9]. The reason behind utilizing the handcrafted data set is to attain performance result when there are no overlapping ADLs as well as failure of event detections. In case of handcrafted data set, we generated transmission patterns of sensors based on the realistic behavioral patterns of the residents in smart home environments, as shown in Figure 3(a) where the horizontal and vertical axes denote time and sensors, respectively. Specifically, the activities that the residents performed throughout a day consisted of 436 real data transmissions. We also have utilized data transmission patterns from [9] as shown in Figure 3(b), which contains 1248 transmissions. For the comprehensive evaluation, we have sampled performance measures 100 times per given fake data ratio values. The confidence interval of clustering accuracy is shown in Figure 4.

4.2. Simulation Results

We have evaluated 3 different privacy protection methods using the performance metrics discussed in the previous section. Throughout the evaluation, we denoted PDT method as probability fitting (prob. fit) since real data transmission is delayed to fit into a specific probabilistic distribution while the immediate transmission method was denoted as *without prob. fit* since real data transmissions are made without delays to fit into the probabilistic distribution. The results illustrate that the magnitude of performance measurements varies depending on the characteristics of given data set. For example, consider Figure 5 that shows the latency of periodic transmission method. When the fake-to-real ratio is 5, Figure 5(a) shows the latency around 1000 seconds while Figure 5(b) shows it around 270 seconds. However, despite the difference in measurement magnitude, the general trends of performance stay similar.

Figure 4 illustrates the relationship between fake-to-real ratio (energy efficiency) and clustering accuracy depending on the protection methods. As expected, the periodic transmission method perfectly protects the privacy of the residents against intelligent adversaries who are utilizing FATS attack. In comparison, PDT method guarantees near perfect privacy where average clustering accuracy is approximately 0.5, independent from the fake-to-real data transmission ratio.² Such independent characteristics of periodic and PDT method from fake-to-real ratio is due to the fact that real data transmissions are delayed to make sure transmission patterns of each sensor is not distinguishable from the other nodes. In contrast, an interesting phenomenon is observed when real data transmissions are made without delays. As shown in the Figure 4(a) and (b), the clustering accuracy values were gradually decreased as the rate of fake data transmissions increased, leading to the increase in privacy preservation level. Based on such observation, we have concluded that number of fake data transmissions are made without delays.

² The clustering accuracy of 0.5 means, on average, a half of the cluster members are correctly chosen, while the rest aren't. This is equivalent to the case of randomly guessing member sensors for each cluster, thus disclosing very little information to the adversary (near perfect privacy).





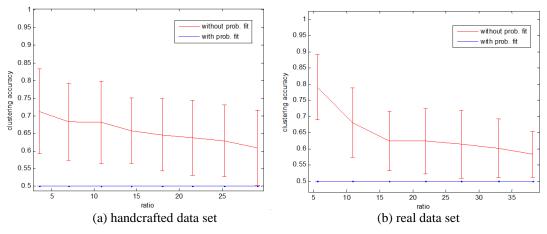


Figure 4. Privacy Results

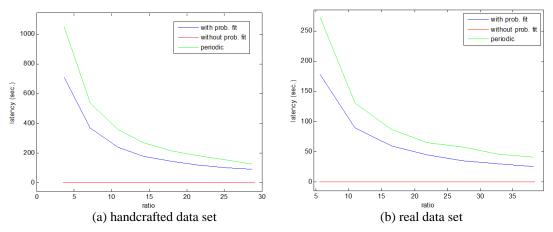


Figure 5. Latency Results

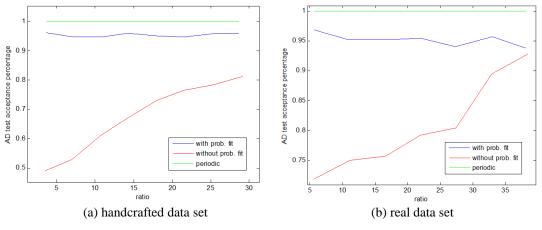




Figure 5 demonstrates the relationship between energy efficiency and latency. For the periodic transmission method, despite its advantage on guaranteeing perfect privacy, the latency is worst compared with other protection methods. Based on the latency trend of periodic method, it was evident that the latency was exponentially decreasing as the number of fake data increased. In case of the PDT method, the latency trend was very similar to the periodic one while the magnitude of latency was approximately 40% better. The latency for immediate transmission method was the best incurring no delays, since real data transmission was made immediately when an event was detected.

Figure 6 illustrates the result of AD test acceptance percentage measured by varying the ratio of fake to real data transmissions. The results demonstrate that the AD test acceptance percentage for immediate transmission method increases as the number of fake data transmissions increases. Such a phenomenon is due to the fact that the capability of obscuring real data transmission patterns increases as more fake entities are added to foster anonymity of interest target. Based on the results shown in Figure 6 (a), we conclude that there must be more than 30 fake data transmission per single real data to provide sufficient degree of AD test acceptance. On the other hand, as shown in Figure 6 (b), approximately 35 fake data transmissions are needed to achieve the acceptance percentage similar to the one provided by PDT method.

5. Conclusion

In this paper, we have study and evaluate the performance of various privacy protection methods. The evaluation results demonstrate that 1) the periodic transmission method was the best choice when perfect privacy is needed while latency is not a significant factor for the performance, 2) the PDT method guarantees near perfect privacy while lessens the latency of real data transmissions compared with the periodic method, and 3) the immediate transmission method can be used if real time event handling is mandated to support optimal QoS while energy efficiency and privacy is not the most significant factors for performance. Based on these evaluation results, we conclude that there is trade-off between privacy, latency, and energy efficiency.

Acknowledgements

This work was supported in part by the IT R&D program of MKE/KEIT [10041145, Self-Organized Software-platform (SOS) for welfare devices], and in part by the DGIST R&D Program of the Ministry of Education, Science, and Technology of Korea (12-BD-0404).

References

- [1] V. Srinivasan, J. Stankovic and K. Whitehouse, "Protecting your daily in-home activity information from a wireless snooping attack", Proceedings of the 10th international conference on Ubiquitous computing, UbiComp, ACM, (2008), New York, USA, pp. 202-211.
- [2] M. Shao, Y. Yang, S. Zhu and G. Cao, "Towards Statistically Strong Source Anonymity for Sensor Networks", Proceedings of the 27th Conference on Computer Communications, INFOCOM, IEEE, (2008), pp. 51-55.
- [3] Y. Yang, M. Shao, S. Zhu, B. Urgaonkar and G. Cao, "Towards event source unobservability with minimum network traffic in sensor networks", Proceedings of the first ACM conference on Wireless network security, WiSec, ACM, (2008), New York, USA, pp. 77-88.
- [4] Y. Zatout, E. Campo and J. Llibre, "WSN-HM: Energy-efficient Wireless Sensor Network for home monitoring", 5th International Conference on Intelligent Sensors, Sensor Networks and Information Processing, ISSNIP, IEEE, (2009), pp. 367-372.
- [5] C. Ozturk, Y. Zhang and W. Trappe, "Source-location privacy in energy-constrained sensor network routing", Proceedings of the 2nd ACM workshop on Security of ad hoc and sensor networks, SASN, ACM, (2004), New York, USA, pp. 88-93.
- [6] United States department of health and human services, "HIPAA regulations and standards", http://www.hhs.gov/orc/hipaa/.
- [7] A. Perrig, J. Stankovic and D. Wagner, "Security in wireless sensor networks", Communications of the ACM, vol. 47, Issue 6, ACM, (2004), pp. 53-57.
- [8] K. Mehta and D. Liu, "Protecting Location Privacy in Sensor Networks against a Global Eavesdropper", IEEE Transactions on Mobile Computing, TMC, vol. 11, no. 2, IEEE, (**2012**), pp. 320-337.
- [9] T. L. M. van Kasteren, A. K. Noulas, G. Englebienne and B. J. A. Kröse, "Accurate Activity Recognition in a Home Setting", ACM Tenth International Conference on Ubiquitous Computing (Ubicomp'08), ACM, (2008), Seoul, South Korea.
- [10] A. Pfitzmann and M. Hansen, "Anonymity, unlinkability, unobservability, pseudonymity and identity management - a consolidated proposal for terminology", Hannes Federrath (Ed.): Designing Privacy Enhancing Technologies; Proc. Workshop on Design Issues in Anonymity and Unobservability; LNCS (2009).

Authors



Homin Park is Ph.D. candidate in the Department of Information and Communication Engineering at Daegu Gyeongbuk Institute of Science and Technology (DGIST). He received the B.S. degree in Computer Science and Systems (CSS) from University of Washington, Tacoma, WA, USA. His research interests include Cyber-Physical Systems (CPS), smart home environments, wireless sensor networks, and intelligent transportation systems.



Taejoon Park is an Associate Professor in the Department of Information and Communication Engineering, Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu, Korea. Prior to joining DGIST, he was an Assistant Professor at Korea Aerospace University, Gyeonggi-do, Korea from 2008 to 2011, a Principal Research Engineer at Samsung Electronics, Gyeonggi-do, Korea from 2005 to 2008, and a Research Engineer at LG Electronics, Seoul, Korea from 1994 to 2000 (promoted to a Senior Research Engineer in 2000). He received the B.S. degree (summa cum laude) in Electrical Engineering from Hongik University, Seoul, Korea in 1992, the M.S. degree in Electrical Engineering from Korea Advanced Institute of Science and Technology (KAIST), Taejon, Korea in 1994, and the Ph.D. degree in Electrical Engineering and Computer Science from University of Michigan, Ann Arbor, MI, USA in 2005. His current research interests are in cyber-physical and networked embedded systems with emphasis on security, reliability, and timeliness. He has authored or coauthored more than 90 papers/patents including essential patents for the DVD standard.



Sang Hyuk Son is Chair of the Department of Information and Communication Engineering at Daegu Gyeongbuk Institute of Science and Technology (DGIST) and DGIST Fellow. He has been a faculty at the Department of Computer Science of University of Virginia and WCU Chair professor at Sogang University. He received the B.S. degree in electronics engineering from Seoul National University, M.S. degree from KAIST, and the Ph.D. in computer science from University of Maryland, College Park. He has served as the Chair of the IEEE Technical Committee on Real-Time Systems, and he is currently serving as an Associate Editor for IEEE Transactions on Computers, Real-Time Systems Journal, and Journal of Computing Science and Engineering. His research interests include real-time systems, database and data services, QoS management, wireless sensor networks, and information security. He has written or co-authored over 270 papers and edited/authored four books in these areas.

International Journal of Smart Home Vol. 7, No. 2, March, 2013