A Study on Privacy and Energy-Efficiency Tradeoff in Smart Home Environments

Homin Park', Taejoon Park', and Sam Chung[2]

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

2 Institute of Technology
University of Washington, Tacoma, WA, USA
chungsa@uw.edu

Abstract. Privacy protection for smart home environments has focused on ensuring event source anonymity by randomly transmitting fake events. Since home devices are increasingly battery-powered, improving obscurity of real events through frequent fake transmissions is not favorable for longevity of the system. By contrast, minimizing the fake transmissions for energy efficiency without understanding the consequences threatens the residents' privacy. We therefore study how to make a *best tradeoff* between the privacy preservation and the energy efficiency in smart home environments. Our simulation results have demonstrated the adversary's chance of successfully inferring private information of residents is less than 10% when the rate of fake transmissions is at least 14 times higher than that of real transmissions, and it approaches 0% with 20-times higher fake transmissions.

Keywords: privacy for smart home environments, energy efficiency, tradeoffs

1 Introduction

Smart home environments, built on top of wireless sensor network to support residents who need automated assistances, are mandated by the government to protect private and personal information, such as medical history and health condition, at all times [1,7]. To meet such regulation, privacy of the individuals living in smart home environments must be protected by ensuring event unobservability and source anonymity, which is usually achieved via fake data transmissions and intentional delays to obscure the sensitive information [2,3,4,6]. Such protection methods guarantees near perfect privacy preservation against intelligent adversaries who are capable of eavesdropping and analyzing ongoing traffic patterns with sophisticated means, such as statistical inference and classification algorithms [1].

Taejoon Park is the corresponding author.

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 National Research Foundation of Korea Grant funded by the Korean Government, under Grant NRF-2011-0004437.

One of the well-known privacy protection methods for smart home environments is a probabilistic distribution-based transmission scheme introduced by Shao *et al.* [2]. The key idea of this scheme is to generate and transmit fake, meaningless packets, the intervals among which are randomly delayed based on a probabilistic distribution, as well as to delay the transmission of real data to ensure complete randomness of the entire traffic. Although this scheme achieves near perfect privacy with minimized latency of real data transmissions, very frequent fake data transmission inevitably causes excessive energy consumption.

The problem of energy efficiency arises from the limited energy budget available for home devices while the energy cost for data transmission is most significant compared to other tasks. Due to the fact that home devices are increasingly battery-powered, improving obscurity of real events through frequent fake data transmissions is not favorable for longevity of the system. By contrast, minimizing the fake data transmissions for energy efficiency without understanding the consequences threatens the residents' privacy. Therefore, acquiring thorough knowledge on the tradeoff relationship between the level of privacy preservation and the frequency of fake data transmissions is very important to satisfy both aspects as desired.

In this paper, we study how to make a tradeoff between the privacy preservation level and the energy efficiency in smart home environments. Assuming the statistical inference algorithm [1] for the attacker and the probabilistic distribution-based scheme [2] for the defender, we derive the notion of privacy preservation level as well as quantify the energy efficiency. We then present the privacy and energy efficiency tradeoff via a comprehensive simulation of realistic activities of daily living for the residents. Our simulation results have demonstrated that (i) the attacker's chance of successfully inferring private information of residents is less than 10% when the rate of fake transmissions is at least 14 times higher than that of real transmissions, and (ii) it approaches 0% (near perfect privacy) with 20-times higher fake transmissions.

The rest of this paper is organized as follows. Section 2 introduces the background of privacy inference and protection for smart home environments. Section 3 presents the tradeoff relationship between privacy preservation and energy efficiency, and Section 4 concludes the paper.

2 Background

The Fingerprint And Timing-based Snooping (FATS) attack introduced by Srinivasan *et al. [1]* is one of the most dangerous privacy inference attacks targeting smart home environments. By accumulating fingerprints and timestamps of transmitted data for a certain period of time, intelligent adversaries are capable of identifying behavioral patterns of the residents regardless of facility layouts and encryption techniques. As shown in Fig. 1, FATS attack utilizes a multi-tier inference algorithm to identify and classify sensor allocations as well as their functionalities, leading to the inference of daily activities of residents. The algorithm (i) detects if residents are at home, away or sleeping (tier-0), (ii) performs clustering of sensors to figure out the number of rooms and residents (tier-1), (iii) classifies the rooms to characterize the room types like bathrooms (tier-2), and finally (iv) classifies the sensors within each room to precisely capture activities like showering or cooking (tier-3).

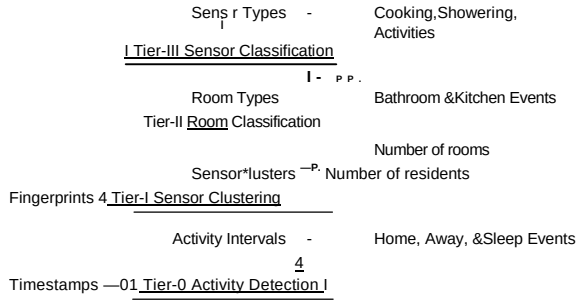


Fig. 1. Multi-tier Approach of a Fingerprints And Timing-based Snooping Attack

A well-known countermeasure against above-mentioned privacy inference attacks is a probabilistic fake data transmission scheme presented by Shao *et al.* [2]. This scheme randomly (i.e., according to a certain probabilistic distribution) transmits fake data traffic as well as delays the real transmission to let outsiders not to tell the real data from fake ones. It minimizes the latency of real data transmission, because the transmission latency is determined by the frequency of fake data transmission, which is controlled by adjusting parameters (e.g., a mean value) of probabilistic (e.g., exponential) distribution. At the same time, it can guarantee near perfect privacy of the residents since adversaries eavesdropping on the ongoing transmissions cannot infer any meaningful information by analyzing the random traffic pattern. Despite these benefits, a serious drawback of the probabilistic fake transmission scheme is the significantly increased energy consumption due to the very frequent fake data transmission required to preserve the level of privacy [4,5].

3 Privacy and Energy Efficiency Tradeoff 3.1

Privacy Preservation Level

To study the tradeoff between privacy and energy efficiency, we derive the *privacy preservation level* from the relevancy of clustering result yielded from tier-1 of FATS attack, based on the key observation that the classification processes in tiers-2 and -3 solely rely on sensor clustering in tier-1 [1]. The basic idea of sensor clustering is that sensors in the same room typically exhibit similar transmission patterns and temporal correlations. Srinivasan *et al.* demonstrated that their tier-1 clustering algorithm is capable of grouping sensors with 80% or higher accuracy regardless of the number of rooms and residents.

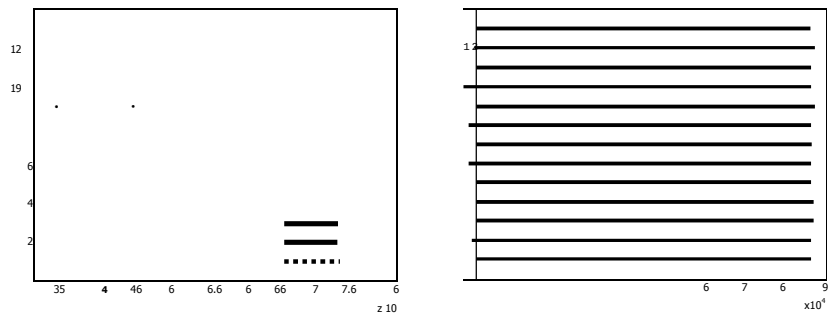
The privacy preservation level has the following physical meaning. If the clustering result is 100% relevant to the actual sensor allocation implying the resident's privacy is perfectly compromised by adversaries, we set its clustering relevancy equal to 1 (the lowest level of privacy preservation). On the other hand, if the clustering result yields no meaningful information related to the actual sensor allocation, we conclude the privacy is perfectly preserved and set the clustering relevancy equal to 0 (the highest level of privacy preservation). Simply put, the privacy preservation level is inversely proportional to the clustering relevancy, i.e., the less relevant the clustered result, the safer the privacy of the resident, and vice versa.

3.2 Energy Efficiency

To quantify the energy efficiency, we utilize a ratio of fake to real data transmissions computed from the number of fake and real data transmissions made. As the *fake-to-real ratio* increases due to the increase in fake data transmissions, the energy efficiency is getting worse because of the extra energy consumed by transferring meaningless information, and vice versa. The reason we utilize the ratio rather than an absolute number of fake data transmissions is to make sure the number of fake data transmissions is greater than real data transmissions.

3.3 Tradeoff Study via Simulation

We present our simulation results to show the *tradeoff* between privacy and energy efficiency. We generated transmission patterns of sensors based on the realistic behavioral patterns of the residents in smart home environments, as shown in Fig. 3 where the horizontal and vertical axes denote time and sensor id, respectively. Specifically, the activities that the resident performed throughout a day consisted of 436 real data transmissions as illustrated in Fig. 2(a). On top of these real data transmission patterns, we generated fake data transmission patterns by applying the exponential distribution as shown in Fig. 2(b). Clearly the change in the privacy preservation level could be observed by varying the parameter, which is a mean value in case of the exponential distribution.



(a) Real Data Traffic (b) Fake Data Traffic

Fig. 2. Real and Fake Transmission Patterns

Fig. 3 shows the relationship between the clustering relevancy (inversely proportional to the privacy preservation level) and the fake-to-real ratio. As shown in the figure, the privacy of residents was perfectly compromised when no fake data transmission occurred throughout a day. The relevancy of clustering result gradually decreased as the rate of fake data transmissions increased, leading to the increase in privacy preservation level. From Fig. 4, we found that the rate of fake data traffic must be at least 14 times higher than the real data traffic to keep the clustering relevancy lower than 0.1. In other words, there must be 6104 fake transmissions when there are 436 real data packets according to the pattern in Fig. 2(a). Furthermore, the fake-to-real ratio must be 20 or higher to meet the need of near perfect privacy (with clustering relevancy close to 0).

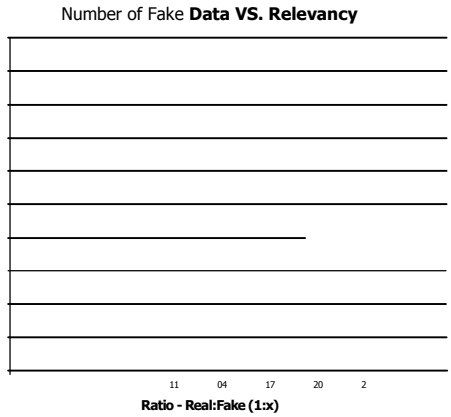


Fig. 3. Tradeoff between Fake-to-real Ratio and Clustering Relevancy

4 Conclusion

In this paper, we have studied the tradeoff between the privacy preservation level (being inversely proportional to the clustering relevancy) and the ratio of fake to real data transmissions. Our simulation results have demonstrated that in order to keep the attacker's chance of successfully inferring private information of residents less than 10%, there must be 14 or more fake data transmissions per a real data generated. Moreover, it approaches 0% (near perfect privacy) with 20-times higher fake transmissions.

References

1. Srinivasan, V., Stankovic, J., Whitehouse, K.: Protecting your daily in-home activity information from a wireless snooping attack. In: Proceedings of the 10th international conference on Ubiquitous computing, UbiComp, pp. 202-211. ACM, New York, USA (2008)
2. Shao, M., Yang, Y., Zhu, S., Cao, G.: Towards Statistically Strong Source Anonymity for Sensor Networks. In: The 27th Conference on Computer Communications, INFOCOM, pp. 51-55, IEEE, (2008)
3. Yang, Y., Shao, M., Zhu, S., Urgaonkar, B., Cao, G.: Towards event source unobservability with minimum network traffic in sensor networks. In: Proceedings of the first ACM conference on Wireless network security, WiSec, pp. 77-88, ACM, New York, USA (2008)
4. Zatout, Y., Campo, E., Llibre, J.: WSN-HM: Energy-efficient Wireless Sensor Network for home monitoring. In: 5th International Conference on Intelligent Sensors, Sensor Networks and Information Processing, ISSNIP, pp. 367-372, IEEE, (2009)
5. Ozturk, C., Zhang, Y., Trappe, W.: Source-location privacy in energy-constrained sensor network routing. In: Proceedings of the 2nd ACM workshop on Security of ad hoc and sensor networks, SASN, pp. 88-93, ACM, New York, USA (2004)
6. United States department of health and human services, HIPAA regulations and standards. http://www.hhs.gov/orc/hipaa/.
7. Perrig, A., Stankovic, J., Wagner, D.: Security in wireless sensor networks. In: Communications of the ACM, vol.47, iss.6, pp. 53-57, ACM, (2004)