Privacy-Aware Platform for Smart Home Sensor Data
（スマートホームセンサデータのプライバシアウェアプラットフォーム）

Name: Sopicha Stirapongsasuti
Laboratory's name: Ubiquitous computing systems
Supervisor's name: Keiichi Yasumoto

Abstract

The development of smart sensors and appliances can provide a lot of services, especially for smart homes, e.g., smart appliance control, anomaly detection, life logging, elderly monitoring, and so on. However, data aggregation from sensors/appliances may contain privacy-sensitive information. Such information may be misused by a malicious attacker who is assumed to be capable of performing re-identification using machine learning techniques since the existing studies show that it is possible to monitor encrypted smart-home data through the network traffic and untrusted cloud. Also, some previous studies attempted to apply privacy mechanisms, but they decreased the utility of services such as activity recognition accuracy. Therefore, it challenges us to achieve these problems. In this thesis, we propose a privacy mechanism to preserve privacy-sensitive sensor data generated in a smart home. The proposed privacy mechanism contains four modules: 1) smart sensors and appliances which generate data in a smart home; 2) home gateways that transfer the data from sensors/appliances to a cloud or edge server; 3) edge servers that process the raw data from sensors/appliances and transfer the processed data to a cloud server; and 4) a cloud server which stores data from home gateways/edge servers and train a machine learning model to generate services to facilitate smart home users. We assume a threat model that an adversary can access some parts or all of the data in the untrusted cloud server at some time slots, while the demands of privacy-protection level from residents differ based on time slot. Thus, we leverage the differential privacy (DP) with feature merging anonymization at edge servers where each edge server has a different privacy budget (ε) to provide privacy by generating synthetic (noise) data. Due to users' different demands and budgets, an optimization technique is used to provide proper decision-making for edge server selection.

To evaluate the trade-off between the activity recognition accuracy and the privacy protection level, we trained the activity recognition and person identification models using a smart-home open dataset with video data for different activities by 15 residents. We applied Gaussian Differential Privacy-based noise to each video frame, extracted HOG features from the frame, and trained the model with the features for different combinations of ε and α values. As a result, we found that the original data showed 71.1% of activity recognition and 92.18% of user identification accuracy, while the activity recognition accuracy lightly decreased to 62.95%, and the user identification accuracy significantly decreased to 57.49% when applying Gaussian DP-based noises. We also conducted a simulation-based study where users have different preferences on privacy protection and activity recognition accuracy depending on time slots and several edge servers with different privacy protection levels are available. As a result, we found that the semi-optimal selection of privacy-protection level for each timeslot, which maximizes users' utility, is possible within the limited resources of edge servers.