IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 3, 2021, accepted September 8, 2021, date of publication September 14, 2021,
date of current version September 29, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3112482

Platform Utilizing Similar Users' Data to Detect Anomalous Operation of Home IoT Without Sharing Private Information

MASAAKI YAMAUCHI, (Graduate Student Member, IEEE), YUICHI OHSITA, (Member, IEEE), AND MASAYUKI MURATA, (Member, IEEE)
Graduate School of Information Science and Technology, Osaka University, Suita, Osaka 565-0871, Japan

Corresponding author: Masaaki Yamauchi (m-yamauchi@ist.osaka-u.ac.jp)

This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number JP21J12993.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Research Ethics Committee of the Graduate School of Information Science and Technology, Osaka University.

ABSTRACT To mitigate the risk of cyberattacks on home IoT devices, we have proposed a method for detecting anomalous operations by learning the behaviors of users based on the operation sequences of their home IoT devices and home conditions. While this method requires a sufficient amount of training data, achieving accurate detection is still possible by utilizing the data of users with similar lifestyles. However, users are unwilling to share their private information with others. In this study, we propose a platform to utilize data of similar users without sharing private information. We introduce an agent that learns behaviors of users to detect anomalous operations in each home and cooperates with other agents. In this framework, an agent requiring cooperation with other agents sends a question to the other agents, attaching identifiers of past questions that are similar to the behaviors learned. The receivers decide whether the question is from a similar agent by using the attached information. If the question is from a similar agent, the agent answers the question. We evaluate our platform by using behavior datasets collected from real homes. We simulate two cases: (1) sequences of operations are monitored, and (2) home IoT devices are used alone but sequences cannot be used for detection. The results show that our framework has a 50.5% higher detection ratio for case (1) when using the behavioral data of each user. For case (2), our framework has a 13.4% higher detection ratio when using all the behavioral data of users.

INDEX TERMS Anomaly detection, cooperative systems, Internet of Things, network security, operation by attackers, secure platform, smart homes.

I. INTRODUCTION

A. MOTIVATION

Consumer electronics such as electric fans and refrigerators have recently been connected to the Internet; these devices are called Internet of Things (IoT) devices. Users can operate these IoT devices by using smartphones and AI speakers via the Internet. Owing to the usefulness of IoT devices, many IoT devices have been installed in homes.

Caused by the popularity of home IoT devices, risks of cyberattacks targeting home IoT devices [2]–[5] and smart homes [6], [7] have increased. Cyberattacks targeting home IoT devices have been previously observed [8], [9]. Currently, these attacks mainly aim to intrude the IoT devices to construct botnets and abuse the IoT devices as step devices for DDoS attacks [10], [11]. Nevertheless, these attacks can be detected by analyzing the behavior of the attacker [12]–[14] and comparing it to the usual behaviors displayed by the home occupants [15], [16].

However, these attacks differ from the attacks that target personal computers and smartphones. This is because home IoT devices are physically close to users [17]. For example, attackers would operate home IoT devices by sending packets via the intruded IoT devices to change the temperature of an air conditioner or unlock a smart home lock. These attacks may make users unsafe and could even cause

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaolong Li.

physical harm. Furthermore, simultaneous attacks on high-power IoT devices can suddenly increase energy demands, which could lead to major power outages [18]. Therefore, the detection and prevention of these attacks are of paramount importance.

To mitigate the risk of attacks on home IoT devices, we proposed a method that could detect attacks that targeted home IoT devices [19]. This method focused on the daily behaviors of the users at home. The method learned their behaviors by studying a sequence of events and the conditions of the home. The observed sequence of events included events on the home network such as the operations of the IoT devices and the entry and exit of the users. The condition of the home was a combination of recorded sensor values when IoT devices were being operated; these include the time of day, temperature, and humidity. When an operation of a home IoT device differed from the learned behavior, this method detected the operation as an anomalous operation.

The simulation results indicated that the method achieved a 95–100% detection ratio of anomalous operations with less than 20% of misdetections when using a sufficient amount of training data. Nonetheless, when we did not have a sufficient amount of training data, the method could not correctly learn the behaviors of the users. In this case, the method recorded false negatives of anomalous operations and continued to do so until a sufficient amount of operations had been monitored.

However, anomalous operations still need to be detected regardless of whether the data amount is sufficient or not. An intuitive approach is to analyze the data of other users by collecting the behavioral datasets of many users [20]. This method uses the privacy information of the users, including the in-home activities of the users; thus, the privacy of the user should be held in high regard. Another approach, which does not rely on the sharing of the behavioral datasets, is to train each model on each dataset and to construct a general model by sharing the learned results [21], [22]. However, this general model constructed via the cooperation between agents may not match the lifestyle of each user. Therefore, this approach cannot achieve accurate detection of anomalous operations.

B. PROBLEM STATEMENT

Owing to the issues mentioned above, a cooperation framework for agents to detect anomalous operations of home IoT devices that satisfy the following requirements is needed.

1) The framework must not use any information to identify the individual users. The framework must not assign any identifiers to the users and agents. In addition, the agents must not share the personal information of the users including the historical behavior of the users and their personal information, such as ages, genders, and jobs.
2) The agents must avoid cooperating with agents of users who have different lifestyles, which may cause inaccurate detections of anomalous operations.

C. CONTRIBUTION AND ORGANIZATION

In this study, we propose a cooperation platform to utilize the data of similar users for anomaly detection of home IoT devices without sharing private information. In this platform, each home has an agent that learns and detects anomalous operations in the home. When an agent cannot decide whether a current operation is legitimate or anomalous, it sends requests to the other agents via the platform. Only similar agents reply to the requests with only one-bit information that is legitimate or anomalous without sharing personal information. In our platform, an identifier is set to a request and is used to judge the similarity. When an agent receives a request that includes behaviors that are similar to the behaviors learned by the agent, the received agent stores the ID of the request. When the agent wants to send a request, it attaches the stored IDs to the request. When an agent receives the request, the agent answers the request that includes IDs that have been stored by the agent. By doing so, the other agents identify the similarity between themselves and the agent sending the request. That is, in our framework, agents can cooperate without sharing the identity of the users. In addition, each agent can choose to answer the request or not. That is, the agents can choose to cooperate with others or not to avoid sharing information that the user may be unwilling to share.

A key idea used in our platform is judging the similarity of each agent from the past answers without using the identifiers of the users. Hence, we can apply our platform to other systems such as shopping recommendation systems.

In summary, our main contributions of this study are as follows:

- We propose a new method to cooperate with similar agents without sharing private information including the identifiers and personal information of the agents and users.
- We simulate the proposed framework for the detection of anomalous operations of home IoT devices.
- We also demonstrate that the cooperation between similar agents by our framework improves the accuracy of the detection of anomalous operations targeting home IoT devices [19].

The rest of this article is organized as follows. We discuss related works, including anomaly detection methods, an anonymous communication method, and cooperative learning methods in Section II. The proposed platform, which does not share private information but utilizes the dataset of similar users to detect anomalous operations via the cooperation of similar users, is described in Section III. Then, we report the evaluation of our framework and the corresponding results in Section IV. Finally, we conclude and discuss possible future work in Section V.

II. RELATED WORK

Related work are discussed in this section. We explain anomaly detection methods of home IoT devices in section II-A. Then we explain an anonymous communication

method, which can be used in our framework, in section II-B. Finally, we explain the methods that train machine learning models via cooperation and discuss the difference between our platform and these methods in section II-C.

A. ANOMALY DETECTION METHOD

Ramapatruni et al. proposed a method to detect anomalous operations by learning user behaviors. This method used Hidden Markov Models (HMM) to learn the normal activities of a user and collected the information obtained from the sensors and/or statuses of the home IoT devices as observations. By using the observations, this method learned the parameters of the HMM. Then, the trained HMM detected an anomalous operation when the probability of that operation occurring was low. They demonstrated the accuracy of this method by using the dataset collected in a smart home environment. This method focused on the case of a single user [23]. However, a smart home may have multiple users.

Therefore, we have proposed a method to detect anomalous operations even in the case of multiple users [19]. This method detects anomalous operations at the home gateway, which is connected to all home IoT devices, home IoT sensors, and smartphones. First, the home gateway collects two kinds of information. One is the condition information of the operations of home IoT devices, such as the time of day, room temperature, and humidity from the connected home sensors. The other information is the presence/absence of the users in the home from the attaching/detaching information of their smartphones. The home gateway subsequently classifies the conditions of the home by constructing a table of sensed values and stores the sequences of operations of IoT devices and the leaving/entering of users in each cell of the condition table. Finally, the home gateway judges whether legitimate or anomalous operations have occurred by comparing sequences of current operations with the stored sequences of the current condition. This method can handle the case with multiple users by constructing the sequences from the monitored operations and considering the case with multiple users.

We have also proposed a method to define the condition of the home for the detection of anomalous operations [24]. In this method, we defined the conditions of the home by the in-home activities. This method modeled the in-home activities of the users as a state transition model. We defined the state of the home as a combination of the state of the users and the state of the devices. The state of the users was defined by the multiple thresholds of sensor values, such as room temperature, noise, and pressures. The state of devices was defined by the time before or after their operation. This method calculated the transition probabilities. After the calculation, when an operation occurred, the method estimated the current condition by using the modeled transition probabilities. By using the estimated current condition, we could detect anomalous operations.

The above methods accurately detected anomalous operations when the amount of training data was sufficient. However, these methods overlooked a large number of anomalous operations until a sufficient amount of operations was monitored.

Therefore, in this study, we propose an anomaly detection platform to utilize the behaviors of similar users to achieve accurate detection, even if there is an insufficient amount of training data.

B. ANONYMOUS COMMUNICATION

In our platform, agents need to hide their sender information, such as IP addresses and user identifiers, when they communicate with other agents. In this study, we used Tor [25] as an anonymization tool. Tor is a famous anonymous communication tool. By communicating via the Tor network, agents can hide their sender IP address information. In the Tor network, only the IP address information of the sender is hidden but the data of the sending packets are not encrypted. Our platform does not need the data file to be encrypted; thus, our platform uses Tor for the anonymization of communications.

C. COLLABORATIVE LEARNING

There are some kinds of learning methods using multiple datasets.

One of the cooperative learning methods generates big data by collecting data from many clients in one place, such as a datacenter [20]. In this method, the collected data is used to train a machine learning model. When a new client that requires the model joins the service, the client receives the trained model. A new client receives one of the models trained by datasets of similar attributions and trains the model by using data collected by the clients. This method is vulnerable to attacks that may target the data center; the collected data may be leaked if the service on the data center is vulnerable. An example is when an attacker steals user data on the Internet cloud through a misconfigured web application firewall [26].

Privacy-preserving machine learning is an approach used to train machine learning models hiding private information [21]. One of the methods of privacy-preserving machine learning is to use differential privacy [27]. The differential privacy preserves the data privacy of users by adding random noise to the data. By collecting a large amount of the noised data, we can train a machine learning model to preserve privacy (the influence of noise can be eliminated statistically).

However, this approach is difficult to apply to the detection method of anomalous operations on home IoT devices. This is because the normal behavior of the users depends on their lifestyle, and accurate private information that can identify the lifestyles is required to achieve accurate detections.

Federated learning is an approach that trains machine learning models by the cooperation between users [22]. In this approach, an agent is deployed for each user. Each agent first trains the model independently by using the data obtained by itself. Then, agents share the trained models and construct the general model by combining the shared models. This approach can train the machine learning models without sharing private information. However, this approach also has

TABLE 1. Comparison of cooperative methods: The characteristics of various existing cooperative methods and the proposed method.

Characteristics	Centralized learning [20]	Centralized learning with differential privacy [27]	Federated learning [22]	Proposed platform
Centralized/distributed	centralized	centralized	distributed	distributed
Shared information	raw data	noised data	gradient of each trained model	yes/no answer
Generalized model/personalized model	personalized	generalized	generalized	personalized
Requirements 1: Non-sharing private information		✓	✓	✓
Requirements 2: Utilizing only similar data	✓			✓

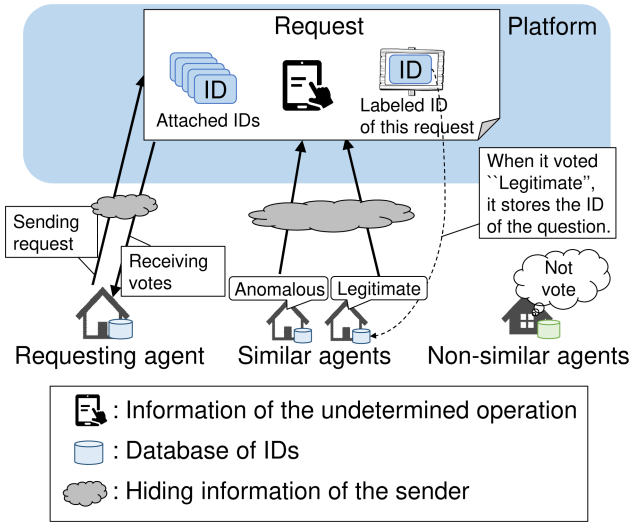


FIGURE 1. Overview of the proposed platform.

difficulty in handling the lifestyles of each user. The general model constructed by the cooperation among all agents may not match the lifestyle of each user. As a result, this approach cannot achieve accurate detection of anomalous operations.

Table 1 shows the summary of the comparison. As shown in this table, any existing approaches do not satisfy the requirements mentioned in Section I-B. Therefore, in this paper, we propose a new framework that does not need the sharing of private information, where only agents with similar data can cooperate.

III. PLATFORM TO UTILIZE SIMILAR DATA OF USERS TO DETECT ANOMALOUS OPERATIONS WITHOUT SHARING PRIVATE INFORMATION

Fig. 1 shows an overview of our platform. In our platform, an agent is deployed for users in a home to learn the behaviors of the users and detect anomalous operations. When an agent cannot decide whether a current operation is legitimate or anomalous, it sends requests to the other agents via the platform to cooperate with them and decide whether the operation is legitimate or anomalous.

Each agent receiving the request first checks whether the request is sent from a similar user. Here, ''similar'' means that the user and the receiver of the request have the same behaviors. If the sender has a similar user, the agent checks whether the behavior of operations included in the request is legitimate or not based on its learned behavior. Then, the agents vote based on their decision. By these steps, the platform collects the votes from similar agents. As a result, the agent that sends the request decides whether the current operation is legitimate or not by checking the results of the votes.

Our platform performs the above steps without identifying any agents. In our platform, the identifiers are set only to the requests. Thus, the other agents cannot identify the origin of the request.

The similarities between agents are calculated based on the IDs of the requests. The sender of a request attaches some IDs of past requests that the sender regards as legitimate to the request. Other agents use the attached IDs to identify if the sender of the request has learned similar behaviors.

A. PROCEDURE OF THE AGENT SENDING REQUEST

Fig. 2 shows the procedure of the requesting agent. When a home gateway detects an operation of a device, an agent of the home checks whether the operation is legitimate or not. If the agent cannot determine whether the operation is legitimate or not due to a lack of learned data, it sends a request to the platform. The request includes the information of the undetermined operation that is used to identify whether the operation is legitimate or not. The sender randomly selects and attaches x number of IDs of the past requests that are identified to be legitimate by the sender agent; the x is a hyperparameter of attaching IDs. This information is used to check whether the sender has learned similar behavior to agents receiving the request. When the hyperparameter x is set to a certain value, the randomness of the selection is not significantly affected by the judgment of the similarity. When sending the request to the platform, the sender can also hide the information of who sent the request from the platform by using tools such as Tor [25]. After sending the request, the sender agent waits for the votes from the other agents.

When the agent receives the votes from the others via our platform, it calculates the number of votes for ''Legitimate''. If the number of votes to ''Legitimate'' is greater than the predefined threshold T, the agent regards the current operation as legitimate.

B. PROCEDURE OF AGENTS RECEIVING REQUEST

Fig. 3 shows a flow chart of the agents who receives a request. When an agent receives a request, the agent first checks the IDs attached to the received request. The agent then compares the attached IDs to the ID database that stores IDs of requests that the receiving agent identifies as legitimate. The similarity

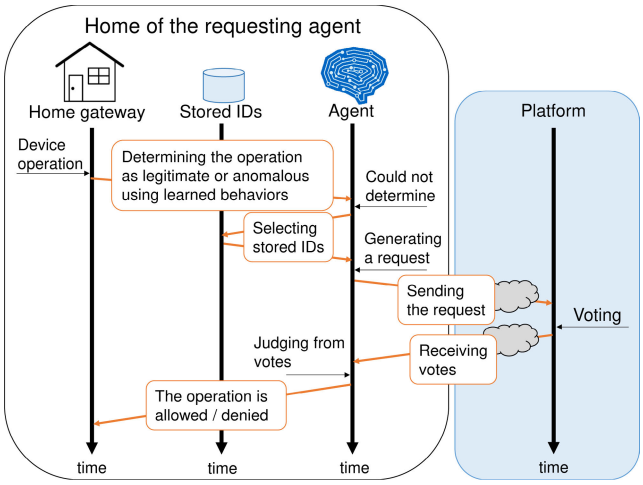


FIGURE 2. Flow chart of the requesting agent.

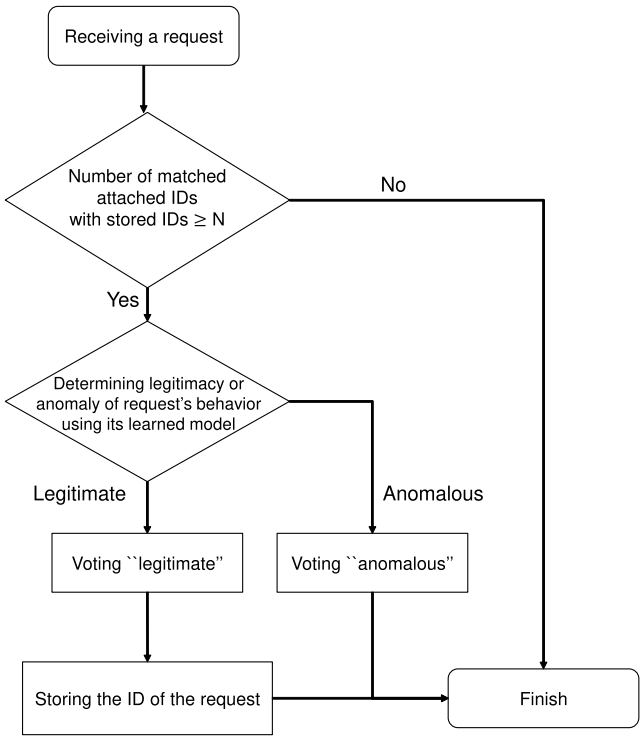


FIGURE 3. Flow chart of the agent receiving a request.

between the sender and receiving agent is judged by the number of matched IDs.

If the number of matched IDs is smaller than a threshold N, the agent does not vote for the request; this is because the sender has different behaviors. By doing so, we avoid the degradation of anomaly detection that could be caused by using data of non-similar users whose behaviors are different. Fig. 4 shows the procedure of the non-similar agents receiving a request.

If the number of matched IDs is larger than the threshold N, the agent judges the sender to be similar to itself.

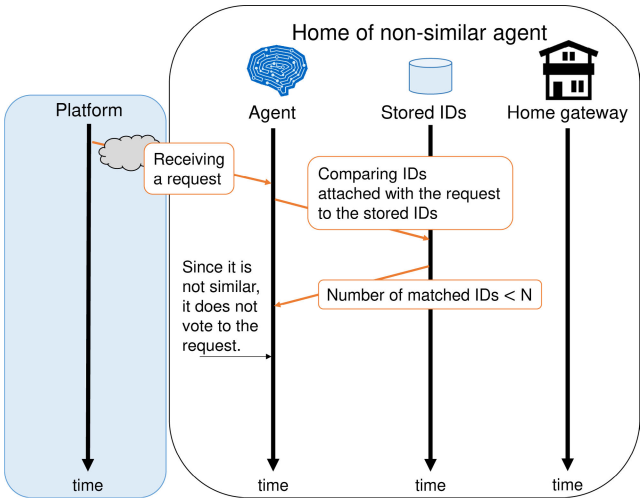


FIGURE 4. Procedure for a non-similar agent receiving a request.

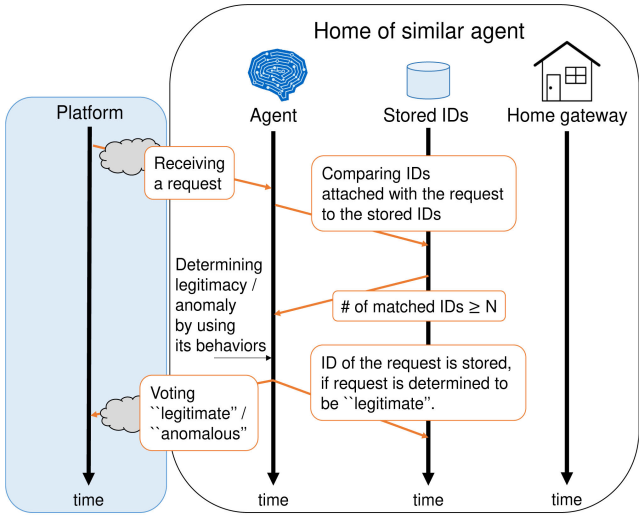


FIGURE 5. Procedure for a similar agent receiving a request.

Subsequently, the agent checks if the behavior included in the request is legitimate or not by using its learned model. Then, it votes by returning its decision to the platform. The decision can either be "Legitimate", "Anomalous", or "Unknown". "Unknown" is a case where the agent does not have a sufficient amount of data to determine whether the behavior included in the request is legitimate or not.

In our platform, even if the number of matched IDs is larger than N, the agent can avoid voting if the user does not want to answer.

When the agent identifies the behavior included in the request as "Legitimate", the agent stores its ID into its ID database. The stored ID is used to identify the similarity of a future request of the agent. As the number of attached IDs to a request becomes larger, receiving agents can estimate the similarity more accurately. Fig. 5 shows the procedure of similar agents receiving a request.

IV. EVALUATION AND RESULTS

We have defined two requirements in Section I-B; cooperation without sending personal information and cooperation with only similar agents. The former is achieved by the proposed platform because the platform does not require users to share their identifiers or their behaviors at home. Therefore, in this section, we demonstrate that our platform satisfies the second requirement. For this evaluation, we implemented and evaluated our platform and compared it to other methods by simulating on datasets captured in real homes. In addition, we checked which agents cooperated to show whether agents cooperated with similar agents.

A. EVALUATION SCENARIO

For this evaluation, we considered the case where a new agent was deployed at a home. The agent learned the behaviors of the users for five days; however, it did not have a sufficient amount of training data. Therefore, the agent cannot detect anomalous operations accurately without cooperating with other agents. Nevertheless, there still are agents that have been deployed at other homes before. These agents have enough data and can detect anomalous operations accurately. Therefore, the new agent would like to join our framework to cooperate with such agents who can detect anomalous operations accurately, regardless of the different lifestyles between the users and the user of the new agent.

To cooperate with the other agent, the newcomer agent needs to store the IDs of the questions similar to the behavior of the corresponding user. One approach to achieving this is to send past requests to the newcomer agents. This situation was evaluated with our method.

By this evaluation, we demonstrate that the cooperation within our framework improves the accuracy of the detection, even if the agent does not have sufficient training data yet.

B. EVALUATION ENVIRONMENT

In this subsection, we describe the settings of our evaluation.

1) DATASET

To evaluate our platform, we collected datasets of activities in two real homes. We recorded the behaviors of the subjects living in the homes, including the time when home appliances were being operated, and the time the subjects entered and left the home. Some home appliances were not connected to the Internet, and we asked the subjects to record the time. To record the logs easily, we installed systems that included buttons, access points, and computers, as shown in Fig. 6. In this system, when a button was pushed, the button sent packets to a computer and the computer recorded the name of the button and the time of day. We put multiple buttons near each home appliance and named the buttons after the name of the home appliance. Their names and the action of the users are shown in Table. 2. We asked the subjects living in the homes to push the button when they used the corresponding consumer electronics or when they left or entered the homes.

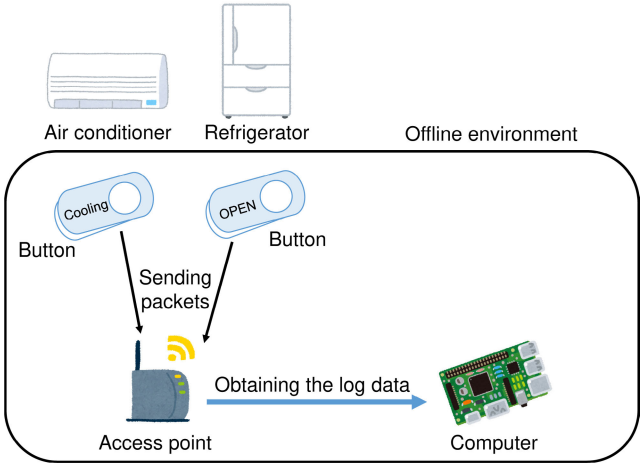


FIGURE 6. System for recording time logs of using home appliances in real homes.

TABLE 2. Collected operations and events by our experimental system deployed in real homes.

Device/event	Action
User position	Entry/exit
Lighting	ON/OFF
Air conditioner	Cooling/heating/dry/raise/lower/OFF
Electric Fan	ON/OFF
Heater	ON/OFF
Washing machine	ON
Refrigerator	Opening
TV	ON/OFF
Cooking stove	ON/OFF
Microwave	ON
Toaster oven	ON
Rice cooker	ON

The collection of data on the in-home activities of users in real homes received approval from the Research Ethics Committee of the Graduate School of Information Science and Technology, Osaka University.

After collecting the logs, we divided the logs into multiple parts so that each part included the monthly data of a home. We collected data for 10 months, from September 2018 to August 2019, for home A. The data were divided and named $A_1, A_2, \ldots, A_{10}$. Data spanning 11 months was collected in home B, from January 2019 to November 2019, and was subsequently divided and named $B_1, B_2, \ldots, B_{11}$. For this evaluation, we simulated the case of multiple homes and some of the users who had have different lifestyles by considering each of A_1 to A_{10} and B_1 to B_{11} to the data of each home. That is, 21 homes participated in our framework in this simulation.

2) ANOMALY DETECTION METHODS APPLIED TO EACH AGENT

In our platform, agents could cooperate with any anomaly detection methods. Each agent independently detected anomalous operations and asked other agents via our platform if the agent could not identify whether an operation was legitimate or anomalous.

For this evaluation, we applied our anomaly detection method [19] to each agent of our platform. This method learned the behaviors of the user as a combination of the condition of the home and the sequences of operations in the home. When users operated home IoT devices continuously, this method utilized the sequence information. According to the results of our previous work, the sequence of the operations plays a significant role in the identification of legitimate operations. However, we need a sufficient number of monitored sequences to train the model on the behavior of the users and achieve accurate detection. If each agent does not have a sufficient number of sequences, the agent alone cannot accurately identify legitimate operations. Thus, our framework, which enables cooperation between agents, is required. For this evaluation, we demonstrate that our platform works well for the agents that use the sequence of operations.

Though the sequence of operations is very powerful, we cannot use sequences that contain only a single operation; that is, the operations on the IoT devices that are used alone. In such a case, our detection method uses only the condition information, such as time of day, to identify the legitimate operations on such devices. The detection based on the condition information also requires a significant amount of the monitored operations that are used to train the model of the behavior of the users. However, the conditions where each IoT device is used depend on the lifestyles of the users. It should be noted that the information of the users whose lifestyles differ may degrade the accuracy of the detection. In this study, we also demonstrate that our platform works in such cases.

Therefore, we simulate two cases; (1) the case where the sequences of operations are monitored, and (2) the case where the device is used alone and the sequence cannot be used.

3) METRICS

For this evaluation, we used two metrics: the detection ratio and the number of misdetected legitimate operations.

We considered the operations by the users included in the dataset to be legitimate operations. The number of misdetected operations was the number of legitimate operations detected as anomalous. For this evaluation, each home included a different number of legitimate operations. Therefore, the effects of one misdetection were different for each home. To ensure that the evaluation of the effectiveness of our method for each home was the same, we did not evaluate using the ratio of misdetections.

The detection ratio was defined by the number of anomalous operations detected as anomalous divided by the number of all inserted anomalous operations. For this evaluation, we inserted 100 anomalous operations per day into the test dataset at random times and calculated the detection ratio. For this evaluation, each home had a dataset with a different number of days. Since each home had a different number of inserted anomalous operations, we compared the detection accuracy using the ratio of detected anomalous operations.

TABLE 3. Details about proposed platform and comparative methods.

Method	Cooperation	Similarity
Cooperate with only similar (proposed)	✓	✓
Do not cooperate		—
Cooperate with all	✓	

The detection ratio of anomalous operations and the number of misdetected operations depended on the parameter of the detection methods and our framework. However, there was a tradeoff; the parameters set to detect more anomalous operations caused more misdetections. Therefore, we changed the parameters of the methods and obtained the ratios of the detected anomalous operations and the number of misdetected operations.

4) COMPARED METHODS

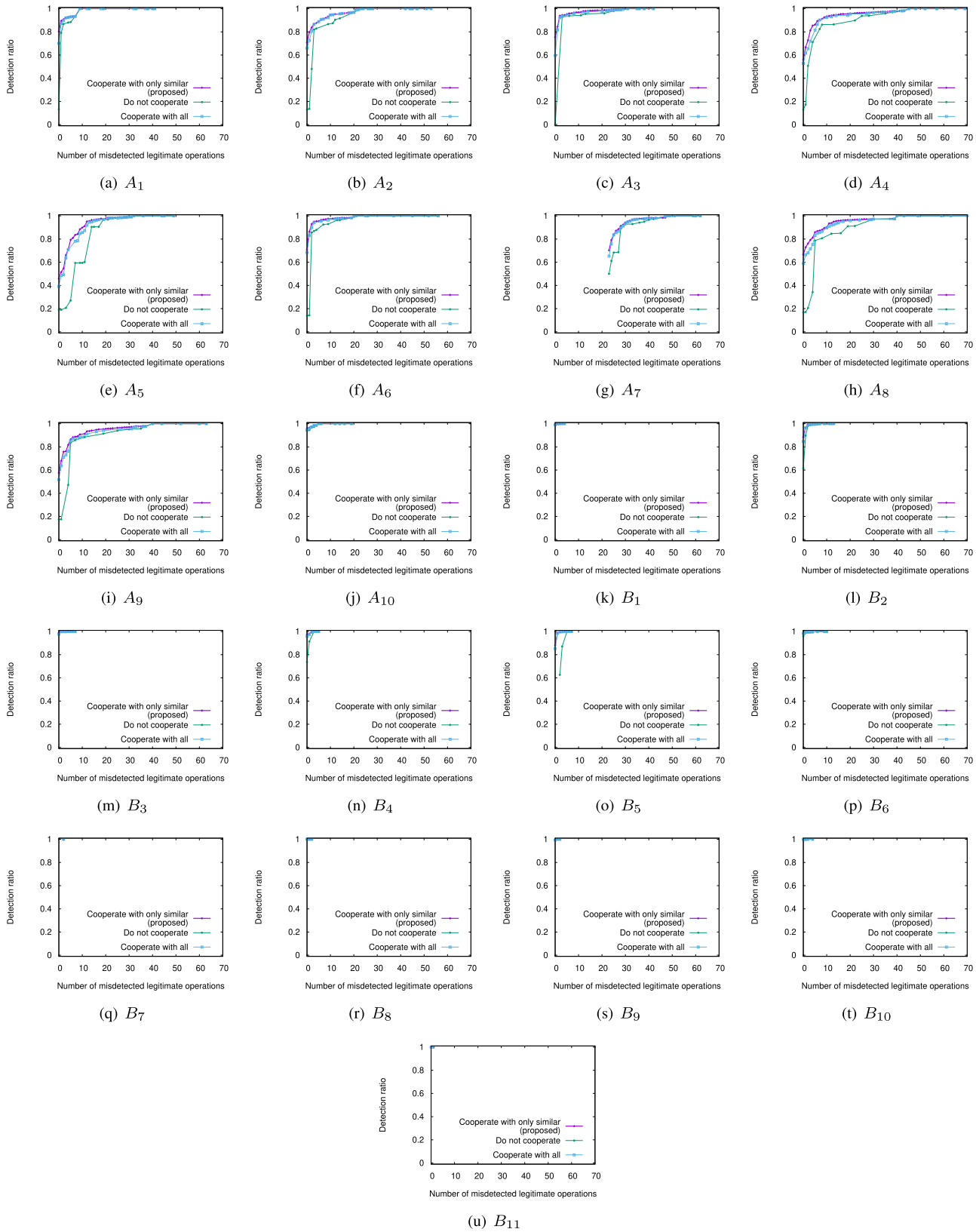
For this evaluation, we compared our cooperation platform with two methods; "Do not cooperate" and "Cooperate with all", as shown in Table. 3. "Do not cooperate" is where each agent does not use our platform but performs detection by only using the behavior data of the users monitored by itself. By comparing the cooperation platform to "Do not cooperate", we demonstrate the effectiveness of cooperation with other agents. "Cooperate with all" is where agents cooperate with all the other agents via our platform. By comparing the cooperation platform to "Cooperate with all", we demonstrate the effectiveness of cooperation with only similar agents.

5) EVALUATION STEPS

We first selected a newcomer agent. Then we divided the one-month data corresponding to the selected agent into two half-month segments. We used the second part as the test data when the first part was used to train the model, and vice versa. By doing so, we can use the full data as the test data. After selecting the newcomer agent and the test data, we trained the model of the newcomer agent by using the data for the first five days that were not included in the test data. The other agents were the agents that cooperated with the newcomer agent and their models were trained using all of the data collected when the corresponding home was being monitored. After the models were trained, we simulated the preparation of the cooperation by sending past requests to new agents. For this simulation, we used the data of the first 14 days as the past requests sent to the agents. Finally, we simulated our framework again by using the test data of the newcomer agent. We evaluated all homes equally and only changed the new agent.

6) PARAMETERS

Our platform has three parameters x, T, and N. We set x to 100, T to $1, 2, 3, \ldots, 20$, and N to $1, 2, 3, \ldots, 100$ for our framework. We set x to 100, T to $1, 2, 3, \ldots, 20$, and N to 0 for the "Cooperate with all" method.



(a) A_1 (b) A_2 (c) A_3 (d) A_4

(e) A_5 (f) A_6 (g) A_7 (h) A_8

(i) A_9 (j) A_{10} (k) B_1 (l) B_2

(m) B_3 (n) B_4 (o) B_5 (p) B_6

(q) B_7 (r) B_8 (s) B_9 (t) B_{10}

(u) B_{11}

FIGURE 7. Detection results of each home. Each agent performed the anomaly detection method using by considering both the condition of the home and the sequence of operations.

TABLE 4. Parameter values of the anomaly detection method applied to agents of each home that learn by combining condition and sequence information.

Parameter	Set values
T	600
α	0, 900, 3600, 10800, 32400, 43200
n_1	0.00, 0.10, 0.25, 0.50, 1.00
$n_d(d \geq 2)$	0.00, 0.10, 0.25, 0.50, 1.00

TABLE 5. Parameter values of the anomaly detection method applied to agents of each home that learn only from the condition information.

Parameter	Set values
T	600
α	0, 150, 300, 450, ..., 1200, 1500, 2100, 3600, 7200, 10800, 18000, 25200, 32400, 43200
n_1	0.05, 0.10, 0.15, ..., 0.95, 1.00
$n_d(d \geq 2)$	1.00

The anomaly detection method used in this evaluation also had parameters T, α, n_1, $n_d(d \geq 2)$, and home condition. For this evaluation, we defined the home condition by using only the time-of-day information. We varied the parameters, as shown in Table 4 and 5.

C. RESULTS

For this evaluation, we added anomalous operation for the cooking stoves and evaluated the accuracy of the detection.

Fig. 7 shows the detection results of the proposed platform and the compared methods when we use the sequences of the operations to detect anomalous operations. In this figure, the horizontal axis represents the number of misdetected legitimate operations and the vertical axis represents the detection ratio. The figure is a plot of the achievable detection ratio against the number of misdetected operations not exceeding the given value on the horizontal axis. These figures indicate that homes A_2, A_4, A_5, A_6, A_7, A_8, A_9, B_1, and B_5 has a reduction in the number of false negatives when using our platform. They achieve higher detection ratios than the non-cooperation method when the parameters are set to ensure that the number of misdetections is less than 20. Specifically, home A_5 records a significant reduction in the number of false negatives. The detection ratio of the proposed method for home A_5 is 50.5% higher than the non-cooperation method when the parameters are set to achieve less than four misdetections; this is less than one misdetection per week. The number of misdetections is greater than 23 for home A_7; this is because one of the half-month data for the home does not include the operation data of the first five days. When we ignore the 23 misdetections, the results show that our platform can reduce the number of false negatives. In other words, cooperation improves the accuracy of the detection. This is because the agents cannot learn the behaviors of the users sufficiently from the data of the first five days. Nevertheless, by using our platform, the agents avoid the case where false negatives of anomalous operations are caused by the lack of training data by cooperating with similar agents.

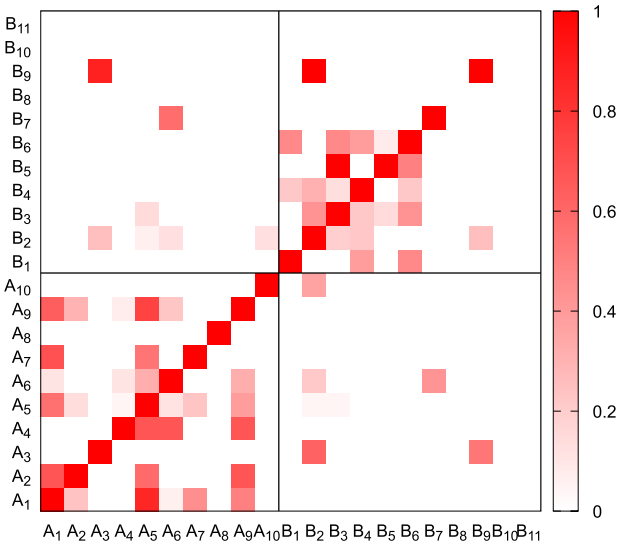
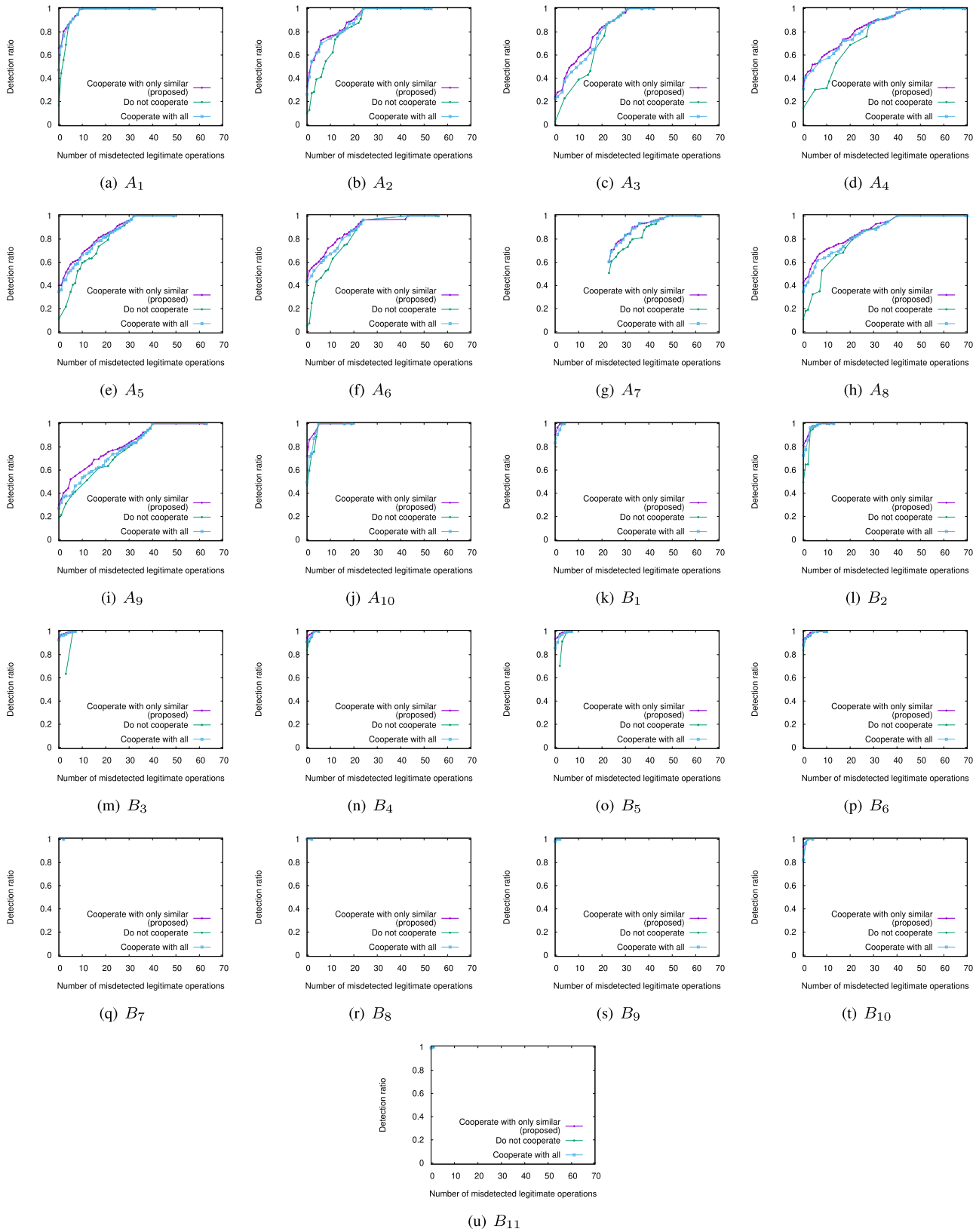


FIGURE 8. The heatmap of matched IDs of each agent; percentages of the ID database on the vertical axis that matched with the ID database of the user on the horizontal axis.

When we have enough training data in each home, which is the case for homes A_1, A_3, A_{10}, B_2, B_3, B_4, and B_6, our framework cannot reduce the number of misdetections. This is also true when the number of legitimate operations is too small to cooperate with other agents, such as for homes B_7, B_8, B_9, B_{10}, and B_{11}. In this case, each agent has only a small amount of operations and few requests are stored as requests that are similar to the behavior of the corresponding users. As a result, the agents cannot cooperate with other agents via our platform. However, the misdetections in such homes are not significant because there are only a small number of operations.

Nevertheless, the detection accuracy of our platform is similar to the method that used cooperation between all agents. This is because the sequences of the operations significantly aid in the detection of anomalous operations. The method using the sequences detects anomalous operations unless the current operations match the sequence of operations, including the operations of the other devices. For this evaluation, we added the anomalous operations on the cooking stoves, and the added operations were rarely included in the sequences that matched the legitimate sequence. As a result, should an agent cooperate with the agents of users who have different behaviors, such cooperation does not degrade the detection ratio.

Fig. 8 shows the heatmap of the similarity between the IDs of the requests stored by each agent. We calculated the similarity of the stored IDs by the percentages of IDs stored by the agents in the vertical axis that matched the IDs stored by the agents in the horizontal axis. For our platform, agents storing the same IDs in their ID database were defined as similar agents. From this figure, agents who corresponded



(a) A_1 (b) A_2 (c) A_3 (d) A_4

(e) A_5 (f) A_6 (g) A_7 (h) A_8

(i) A_9 (j) A_{10} (k) B_1 (l) B_2

(m) B_3 (n) B_4 (o) B_5 (p) B_6

(q) B_7 (r) B_8 (s) B_9 (t) B_{10}

(u) B_{11}

FIGURE 9. Detection results of each home. Each agent performed the anomaly detection method using the condition of the home.

to the same real home cooperated in our platform. That is, our platform achieved cooperation with agents with similar behaviors.

We also evaluated our framework when the anomaly detection method was only based on the condition of the home. Fig. 9 shows the results for each agent, where the anomaly detection method is solely based on the condition of the home. We plotted the results in the same way as Fig. 7. These figures indicate that homes A_3, A_4, A_5, A_6, A_8, A_9, A_{10}, B_1, B_2, and B_5 achieve a smaller number of false negatives of anomalous operations than the method that has cooperation between all agents. When the parameters are set to achieve a misdetections number less than four, the detection ratio of the proposed method is 13.4% higher than that of the method that has cooperation between all agents in home A_4. This is because cooperation with agents of different behaviors degrades the accuracy of the detection. By cooperating with all agents, an agent uses information that does not match the behavior of the corresponding users. As a result, some attacks that are different from the behavior of the corresponding users but match the behavior of other users are not detected. However, in our framework, agents cooperate with only similar agents. As a result, our framework avoids the degradation of the detection accuracy that is caused by using the information of users whose lifestyles are different.

Similar to Fig. 7, our method cannot reduce the misdetections for the case where the number of legitimate operations is too small to cooperate with other agents; such is the case for homes B_7, B_8, B_9, B_{10}, and B_{11}. However, if an agent has some legitimate operations and can calculate the similarities between agents, our framework achieves a smaller number of misdetections than the method that incorporates cooperation between all agents.

V. CONCLUSION AND FUTURE WORK

In this study, we proposed a cooperation framework that utilized the dataset of similar users without sharing their private information to detect anomalous operations. In this platform, an agent was deployed at each home. The agent learned the behavior of the users and detected anomalous operations based on the learned behaviors. However, if the agent did not identify whether the current operation was legitimate or not, due to a lack of training data, it asked the other agents by sending a request. The agent informed the property of the users by attaching the IDs of the past requests that matched the behavior of the corresponding users. Then agents who also identified the past requests of the attached IDs as legitimate replied to it. By doing so, our framework enabled agents to cooperate with similar agents without sharing their private information.

We evaluated our framework by using the dataset monitored at real homes. The results indicated that our framework reduced the number of false negatives of anomalous operations by cooperating between similar agents.

We applied our framework to the detection of anomalous operations of home IoT devices. However, our framework can also be applied to other scenarios where cooperation between similar users is required; this can serve as future work. For example, by applying our framework to a system that recommends products to users, an agent recommends a product using the information on similar users without sharing their private information.

In this study, we assumed that all agents behaved correctly. However, we should consider the case where attackers join our framework to evaluate its robustness. The defense mechanism against such attacks targeting our framework is another future work.

ACKNOWLEDGMENT

This paper is an extension of our previous paper [1] [DOI: 10.1109/ICCE-Taiwan49838.2020.9258353]; it investigates and discusses the effectiveness of our method and compares it to methods that use a part of our learning method.

REFERENCES

[1] M. Yamauchi, Y. Ohsita, and M. Murata, "Platform utilizing others' behavior data to detect anomalous operation hiding private information," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Taoyuan, Taiwan, Sep. 2020, pp. 1–2.

[2] I. Lee and K. Lee, "The Internet of Things (IoT): Applications, investments, and challenges for enterprises," *Bus. Horizons*, vol. 58, no. 4, pp. 431–440, 2015.

[3] M. U. Farooq, M. Waseem, A. Khairi, and S. Mazhar, "A critical analysis on the security concerns of Internet of Things (IoT)," *Int. J. Comput. Appl.*, vol. 111, no. 7, pp. 1–6, 2015.

[4] B. L. R. Stojkoska and K. V. Trivodaliev, "A review of Internet of Things for smart home: Challenges and solutions," *J. Cleaner Prod.*, vol. 140, no. 3, pp. 1454–1464, 2017.

[5] M. Capellupo, J. Liranzo, M. Z. A. Bhuiyan, T. Hayajneh, and G. Wang, "Security and attack vector analysis of IoT devices," in *Proc. Int. Conf. Secur., Privacy Anonymity Comput., Commun. Storage, (SpaCCS) Int. Workshops*, Guangzhou, China, Dec. 2017, pp. 593–606.

[6] D. K. Madhugundu, F. Ahmed, and B. Roy, "A survey on security issues and challenges in IoT based smart home," in *Proc. 3rd Int. Conf. Internet Things Connected Technol. (ICIoTCT)*, 2018, pp. 423–427.

[7] V. Moustaka, Z. Theodosiou, A. Vakali, and A. Kounoudes, "Smart cities at risk!: Privacy and security borderlines from social networking in cities," in *Proc. Int. World Wide Web Conf.*, 2018, pp. 905–910.

[8] M. Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric, J. A. Halderman, L. Invernizzi, M. Kallitsis, D. Kumar, C. Lever, Z. Ma, J. Mason, D. Menscher, C. Seaman, N. Sullivan, K. Thomas, and Y. Zhou, "Understanding the Mirai botnet," in *Proc. 26th USENIX Secur. Symp.*, Vancouver, BC, Canada, Aug. 2017, pp. 1093–1110. [Online]. Available: https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/antonakakis

[9] D. Palmer. (May 2017). *120,000 IoT Cameras Vulnerable to New Persirai Botnet Say Researchers*. [Online]. Available: https://www.zdnet.com/article/120000-iot-cameras-vulnerable-to-new-pers%irai-botnet-say-researchers

[10] Y. M. P. Pa, S. Suzuki, K. Yoshioka, T. Matsumoto, T. Kasama, and C. Rossow, "IoTPOT: A novel honeypot for revealing current IoT threats," *J. Inf. Process.*, vol. 24, no. 3, pp. 522–533, 2016.

[11] M. Lyu, D. Sherratt, A. Sivanathan, H. H. Gharakheili, A. Radford, and V. Sivaraman, "Quantifying the reflective DDoS attack capability of household IoT devices," in *Proc. 10th ACM Conf. Secur. Privacy Wireless Mobile Netw.*, New York, NY, USA, Jul. 2017, pp. 46–51.

[12] V. Martin, Q. Cao, and T. Benson, "Fending off IoT-hunting attacks at home networks," in *Proc. 2nd Workshop Cloud-Assisted Netw. (CAN)*, 2017, pp. 67–72.

[13] K. Xu, F. Wang, and X. Jia, "Secure the internet, one home at a time," *Secur. Commun. Netw.*, vol. 9, no. 16, pp. 3821–3832, Jul. 2016.

[14] S. Shirali-Shahreza and Y. Ganjali, "Protecting home user devices with an SDN-based firewall," *IEEE Trans. Consum. Electron.*, vol. 64, no. 1, pp. 92–100, Feb. 2018.

[15] K. Xu, F. Wang, R. Egli, A. Fives, R. Howell, and O. Mcintyre, "Object-oriented big data security analytics: A case study on home network traffic," in *Proc. Int. Conf. Wireless Algorithms, Syst., Appl.* Harbin, China: Springer, Jun. 2014, pp. 313–323.
[16] K. Xu, F. Wang, L. Gu, J. Gao, and Y. Jin, "Characterizing home network traffic: An inside view," *Pers. Ubiquitous Comput.*, vol. 18, no. 4, pp. 967–975, Apr. 2014.
[17] N. Komninos, E. Philippou, and A. Pitsillides, "Survey in smart grid and smart home security: Issues, challenges and countermeasures," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1933–1954, 4th Quart., 2014.
[18] S. Soltan, P. Mittal, and H. V. Poor, "BlackIoT: IoT botnet of high wattage devices can disrupt the power grid," in *Proc. 27th USENIX Secur. Symp.*, Baltimore, MD, USA: USENIX Association, Aug. 2018, pp. 15–32.
[19] M. Yamauchi, Y. Ohsita, M. Murata, K. Ueda, and Y. Kato, "Anomaly detection in smart home operation from user behaviors and home conditions," *IEEE Trans. Consum. Electron.*, vol. 66, no. 2, pp. 183–192, May 2020.
[20] J. Qiu, Q. Wu, G. Ding, Y. Xu, and S. Feng, "A survey of machine learning for big data processing," *EURASIP J. Adv. Signal Process.*, vol. 67, pp. 1–16, May 2016.
[21] M. Al-Rubaie and J. M. Chang, "Privacy-preserving machine learning: Threats and solutions," *IEEE Secur. Privacy*, vol. 17, no. 2, pp. 49–58, Mar. 2019.
[22] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 1–19, Feb. 2019.
[23] S. Ramapatruni, S. N. Narayanan, S. Mittal, A. Joshi, and K. Joshi, "Anomaly detection models for smart home security," in *Proc. IEEE 5th Int. Conf. Big Data Secur. Cloud (BigDataSecurity), Int. Conf. High Perform. Smart Comput. (HPSC), IEEE Int. Conf. Intell. Data Secur. (IDS)*, May 2019, pp. 19–24.
[24] M. Yamauchi, M. Tanaka, Y. Ohsita, M. Murata, K. Ueda, and Y. Kato, "Modeling home IoT traffic using users' in-home activities for detection of anomalous operations," in *Proc. 32nd Int. Teletraffic Congr.-PhD Workshop*, Sep. 2020, pp. 1–2.
[25] R. Dingledine, N. Mathewson, and P. Syverson, "Tor: The second-generation onion router," in *Proc. 13th USENIX Secur. Symp.* San Diego, CA, USA: USENIX Association, Aug. 2004, pp. 303–320.
[26] *Seattle Tech Worker Arrested for Data Theft Involving Large Financial Services Company Usao-Wdwa Department of Justice*. Accessed: Mar. 12, 2021. [Online]. Available: https://www.justice.gov/usao-wdwa/pr/seattle-tech-worker-arrested-data-theft-involving-large-financial-services-company
[27] M. U. Hassan, M. H. Rehmani, and J. Chen, "Differential privacy techniques for cyber physical systems: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 1, pp. 746–789, 1st Quart., 2020.

MASAAKI YAMAUCHI (Graduate Student Member, IEEE) received the B.E. and M.E. degrees in information science and technology from Osaka University, Japan, in 2017 and 2019, respectively, where he is currently pursuing the Ph.D. degree with the Graduate School of Information Science and Technology.

His research interest includes network security.

YUICHI OHSITA (Member, IEEE) received the M.E. and Ph.D. degrees in information science and technology from Osaka University, Japan, in 2005 and 2008, respectively.

From April 2006 to March 2012, he was an Assistant Professor at the Graduate School of Economics, Osaka University. In April 2012, he moved to the Graduate School of Information Science and Technology, Osaka University, where he has been an Associate Professor at the Institute for Open and Transdisciplinary Research Initiatives, since January 2019. His research interests include traffic engineering, traffic prediction, and network security.

Dr. Ohsita is a member of IEICE and the Association for Computing Machinery (ACM).

MASAYUKI MURATA (Member, IEEE) received the M.E. and D.E. degrees in information and computer science from Osaka University, Japan, in 1984 and 1988, respectively.

In April 1984, he joined Tokyo Research Laboratory, IBM Japan, as a Researcher. From September 1987 to January 1989, he was an Assistant Professor at the Computation Center, Osaka University. In February 1989, he moved to the Department of Information and Computer Sciences, Faculty of Engineering Science, Osaka University, where he became a Professor of the Graduate School of Engineering Science, in April 1999, and has been with the Graduate School of Information Science and Technology, since April 2004. His research interests include information network architecture, performance modeling, and evaluation.

Prof. Murata is a member of ACM and IEICE.