


Platform Utilizing Others' Behavior Data to Detect Anomalous Operation Hiding Private Information

Masaaki Yamauchi¹,
 Yuichi Ohsita¹,
 Masayuki Murata¹

¹Graduate School of Information Science and Technology, Osaka University.

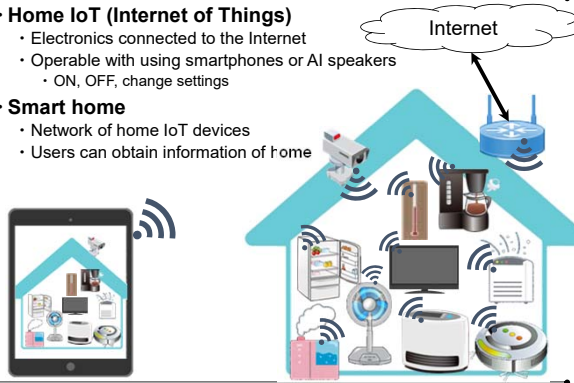


OSAKA UNIVERSITY

September 9th, 2020 IEEE 2020 ICCE-TW - Session H2

Home IoT, smart home


- Home IoT (Internet of Things)**
 - Electronics connected to the Internet
 - Operable with using smartphones or AI speakers
 - ON, OFF, change settings
- Smart home**
 - Network of home IoT devices
 - Users can obtain information of home



September 9th, 2020 IEEE 2020 ICCE-TW - Session H2 1

Anomalous operations of home IoT

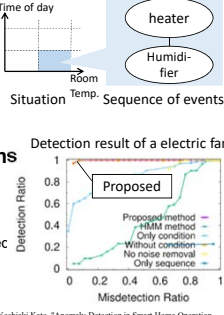
- Attackers send operation packets to home IoT devices**
 - Make users unsafe and may even harm them
 - Operating heater causes burn
 - Change settings of healthcare devices may harm users
- Difficult to detect attacks by the pattern matching**
 - Sending same packets as sent by legitimate users
 - Sending packets via compromised smartphones of legitimate user



September 9th, 2020 IEEE 2020 ICCE-TW - Session H2 2

Detecting anomalous operations [1]

- Modeling user's behaviors as sequences of events for each condition**
 - Sequence of events: order of IoT device operations, users' entering / leaving
 - Condition: time of day and observable sensor values (room temp., noise, ...)
 - Detecting unmatched sequences of operations as anomalous operations with learned behaviors
- Detected 90% anomalous operations with 10% misdetections**
 - Evaluation environment:
 - Installed multiple IoT devices in our lab.
 - 100(/day) anomalous operations were mixed
 - Requiring data for 3 months
 - Less data causes inaccurate detection**

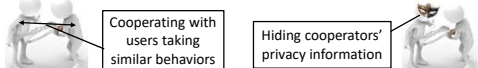


[1] Masaaki Yamauchi, Yuichi Ohsita, Masayuki Murata, Kenzuke Ueda, and Yoshiaki Kato, "Anomaly Detection in Smart Home Operation from User Behaviors and Home Conditions," IEEE Transactions on Consumer Electronics, vol. 66, no. 2, pp. 181-192, May 2020.

September 9th, 2020 IEEE 2020 ICCE-TW - Session H2 3

Problem definition and approach

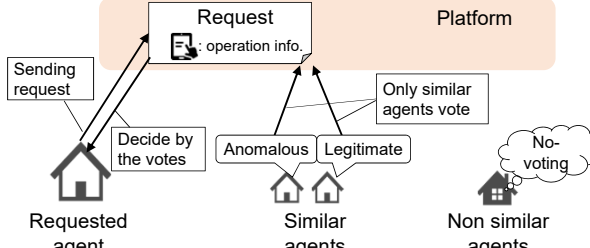
- Inaccurate detection was caused with insufficient training data**
- To use the data collected from similar users**
 - But users do not want to share their private information
 - Private information: behavior history, age, gender, ...
- We propose anomaly detection platform utilizing similar users' data without sharing their private information**
 - [Requirements]
 - Collaborators do not send raw behavior history data
 - Hiding information about data sender (e.g., IP address, User ID)
 - Cooperating with similar users who have same behaviors



September 9th, 2020 IEEE 2020 ICCE-TW - Session H2 4

Proposed platform - base overview

- An agent is deployed to learn behaviors for a home users**
- Sending request to require voting with information of the operation**
 - The operation that an agent cannot decide legitimate or anomalous is sent
 - Only similar agents vote that the operation is "legitimate", "anomalous", "unknown"
 - Similar agents: agents that voted "legitimate" on the same request
 - The requested agent decides legitimate / anomalous based on the votes



September 9th, 2020 IEEE 2020 ICCE-TW - Session H2 5

Problem definition and approach

- Inaccurate detection was caused with insufficient training data
- To use the data collected from similar users
 - But users do not want to share their private information
 - Private information: behavior history, age, gender, ...
- We propose anomaly detection platform utilizing similar users' data without sharing their private information
 - [Requirements]
 - Collaborators do not send raw behavior history data
 - Hiding information about data sender (e.g., IP address, User ID)
 - Cooperating with similar users who have same behaviors

September 9th, 2020 IEEE 2020 ICCE-TW - Session H2 6

Proposed platform - hiding sender information

- Hiding sender's information by using anonymous network
 - e.g., Tor [2] : hiding sender's IP address information
- Identify the similarity by attached IDs to the request
 - When an agent judged a request as legitimate it stores the ID of the request
 - Agents send requests attaching some of the stored IDs
 - Agent votes on the request that includes same IDs as IDs the agent stored

September 9th, 2020 IEEE 2020 ICCE-TW - Session H2 7

Problem definition and approach

- Inaccurate detection was caused with insufficient training data
- To use the data collected from similar users
 - But users do not want to share their private information
 - Private information: behavior history, age, gender, ...
- We propose anomaly detection platform utilizing similar users' data without sharing their private information
 - [Requirements]
 - Collaborators do not send raw behavior history data
 - Hiding information about data sender (e.g., IP address, User ID)
 - Cooperating with similar users who have same behaviors

September 9th, 2020 IEEE 2020 ICCE-TW - Session H2 8

Proposed platform - hiding sender information

- Hiding sender's information by using anonymous network
 - e.g., Tor [2] : hiding sender's IP address information
- Identify the similarity by attached IDs to the request
 - When an agent judged a request as legitimate it stores the ID of the request
 - Agents send requests attaching some of the stored IDs
 - Agent votes on the request that includes same IDs as IDs the agent stored

September 9th, 2020 IEEE 2020 ICCE-TW - Session H2 9

Proposed platform - identify the similarity

- Hiding sender's information by using anonymous network
 - e.g., Tor [2] : hiding sender's IP address information
- Identify the similarity by attached IDs to the request
 - When an agent judged a request as legitimate it stores the ID of the request
 - Agents send requests attaching some of the stored IDs
 - Agent votes on the request that includes same IDs as IDs the agent stored

September 9th, 2020 IEEE 2020 ICCE-TW - Session H2 10

Proposed platform - identify the similarity

- Hiding sender's information by using anonymous network
 - e.g., Tor [2] : hiding sender's IP address information
- Identify the similarity by attached IDs to the request
 - When an agent judged a request as legitimate it stores the ID of the request
 - Agents send requests attaching some of the stored IDs
 - Agent votes on the request that includes same IDs as IDs the agent stored

September 9th, 2020 IEEE 2020 ICCE-TW - Session H2 11

Proposed platform - identify the similarity

- Hiding sender's information by using anonymous network
 - e.g., Tor [2] : hiding sender's IP address information
- Identify the similarity by attached IDs to the request
 - When an agent judged a request as legitimate it stores the ID of the request
 - Agents send requests attaching some of the stored IDs
 - Agent votes on the request that includes same IDs as IDs the agent stored

September 9th, 2020 IEEE 2020 ICCE-TW - Session H2 12

Proposed platform - identify the similarity

- Hiding sender's information by using anonymous network
 - e.g., Tor [2] : hiding sender's IP address information
- Identify the similarity by attached IDs to the request
 - When an agent judged a request as legitimate it stores the ID of the request
 - Agents send requests attaching some of the stored IDs
 - Agent votes on the request that includes same IDs as IDs the agent stored

September 9th, 2020 IEEE 2020 ICCE-TW - Session H2 13

Problem definition and approach

- Inaccurate detection was caused with insufficient training data
- To use the data collected from similar users
 - But users do not want to share their private information
 - Private information: behavior history, age, gender, ...
- We propose anomaly detection platform utilizing similar users' data without sharing their private information
 - [Requirements]
 - Collaborators do not send raw behavior history data
 - Hiding information about data sender (e.g., IP address, User ID)
 - Cooperating with similar users who have same behaviors

September 9th, 2020 IEEE 2020 ICCE-TW - Session H2 14

Proposed platform - hiding sender information

- Hiding sender's information by using anonymous network
 - e.g., Tor [2] : hiding sender's IP address information
- Identify the similarity by attached IDs to the request
 - When an agent judged a request as legitimate it stores the ID of the request
 - Agents send requests attaching some of the stored IDs
 - Agent votes on the request that includes same IDs as IDs the agent stored

September 9th, 2020 IEEE 2020 ICCE-TW - Session H2 15

Evaluation environment

- Constructed a network of home IoT devices in our lab.
 - 15 kinds of consumer electronics and sensors
 - Electric fans, a coffee maker, temperature sensors, etc.
 - 4 subjects
- Captured all packets and sensor data
 - Recorded the times of
 - Operation of home IoT devices
 - Entering or leaving of users
- Divided the collected data into 28 datasets
 - each dataset has one user's operations at 1 month

September 9th, 2020 IEEE 2020 ICCE-TW - Session H2 16

Evaluation

- Dataset
 - Legitimate operations: subjects' operations in the captured data
 - Anomalous operations: added 100 (/day) operations at random time
- Method
 - Each agent learns from 1 user's dataset
 - Cross validation
 - Test data: one half of the agent's dataset added 100 operations
 - Learning data: the other half
 - Sum up each result and calculate *Misdetction*, *False negative ratio*
- Metrics
 - $Misdetction\ ratio = \frac{\# of\ misdetcted\ legitimate\ operations}{\# of\ legitimate\ operations}$
 - $False\ negative\ ratio = \frac{\# of\ not-detected\ anomalous\ operations}{\# of\ added\ anomalous\ operations}$

September 9th, 2020 IEEE 2020 ICCE-TW - Session H2 17

Result

- **Our platform reduced 5 misdetections per 1 month**
 - Our platform could efficiently use other agents' information
- **Cooperating with all agents caused false negatives while it reduced misdetections**
- **Cooperating with only similar agents**
 - reduced misdetections compared with the "single agent"
 - avoided false negatives compared with cooperating with all agents

Sum of 16 users' results about detecting anomalous operation of fans	Misdetected ratio	Misdetected / Total	False negative ratio	False negatives / Total
Single agent	23.6%	30/127	0.493%	224/45176
Platform (cooperating with only similar agents)	19.7%	25/127	0.645%	293/45176
Platform (cooperating with all users)	16.5%	21/127	0.819%	372/45176

September 9th, 2020 IEEE 2020 ICCE-TW - Session H2 18

Result

- **Our platform reduced 5 misdetections per 1 month**
 - Our platform could efficiently use other agents' information
- **Cooperating with all agents caused false negatives while it reduced misdetections**
- **Cooperating with only similar agents**
 - reduced misdetections compared with the "single agent"
 - avoided false negatives compared with cooperating with all agents

Sum of 16 users' results about detecting anomalous operation of fans	Misdetected ratio	Misdetected / Total	False negative ratio	False negatives / Total
Single agent	23.6%	30/127	0.493%	224/45176
Platform (cooperating with only similar agents)	19.7%	25/127	0.645%	293/45176
Platform (cooperating with all users)	16.5%	21/127	0.819%	372/45176

September 9th, 2020 IEEE 2020 ICCE-TW - Session H2 18

Result

- **Our platform reduced 5 misdetections per 1 month**
 - Our platform could efficiently use other agents' information
- **Cooperating with all agents caused false negatives while it reduced misdetections**
- **Cooperating with only similar agents**
 - reduced misdetections compared with the "single agent"
 - avoided false negatives compared with cooperating with all agents

Sum of 16 users' results about detecting anomalous operation of fans	Misdetected ratio	Misdetected / Total	False negative ratio	False negatives / Total
Single agent	23.6%	30/127	0.493%	224/45176
Platform (cooperating with only similar agents)	19.7%	25/127	0.645%	293/45176
Platform (cooperating with all users)	16.5%	21/127	0.819%	372/45176

September 9th, 2020 IEEE 2020 ICCE-TW - Session H2 18

Result

- **Our platform reduced 5 misdetections per 1 month**
 - Our platform could efficiently use other agents' information
- **Cooperating with all agents caused false negatives while it reduced misdetections**
- **Cooperating with only similar agents**
 - reduced misdetections compared with the "single agent"
 - avoided false negatives compared with cooperating with all agents

Sum of 16 users' results about detecting anomalous operation of fans	Misdetected ratio	Misdetected / Total	False negative ratio	False negatives / Total
Single agent	23.6%	30/127	0.493%	224/45176
Platform (cooperating with only similar agents)	19.7%	25/127	0.645%	293/45176
Platform (cooperating with all users)	16.5%	21/127	0.819%	372/45176

September 9th, 2020 IEEE 2020 ICCE-TW - Session H2 18

Result

- **Our platform reduced 5 misdetections per 1 month**
 - Our platform could efficiently use other agents' information
- **Cooperating with all agents caused false negatives while it reduced misdetections**
- **Cooperating with only similar agents**
 - reduced misdetections compared with the "single agent"
 - avoided false negatives compared with cooperating with all agents

Sum of 16 users' results about detecting anomalous operation of fans	Misdetected ratio	Misdetected / Total	False negative ratio	False negatives / Total
Single agent	23.6%	30/127	0.493%	224/45176
Platform (cooperating with only similar agents)	19.7%	25/127	0.645%	293/45176
Platform (cooperating with all users)	16.5%	21/127	0.819%	372/45176

September 9th, 2020 IEEE 2020 ICCE-TW - Session H2 18

Conclusion and future work

- **Anomaly detection platform that utilizes the dataset of similar users without sharing their private information**
- **Useful to reduce the misdetections**
- **[Future work]**
 - To evaluate our method with more dataset
 - In this paper, we used a small amount of dataset
 - The advantages of our platform becomes large as the number of homes increases
 - Countermeasures for attacking the platform
 - In this paper, we assumed that all agents performs correct operations

September 9th, 2020 IEEE 2020 ICCE-TW - Session H2 19