Identification of IoT User Actions in Encrypted Traffic

Pierre-Marie Junges[2], Jérôme François[1], Olivier Festor[2]

Lorraine Research Laboratory in Computer Science and its Applications
(CNRS/[1]INRIA/University of Lorraine[2])

May 16, 2019

Introduction

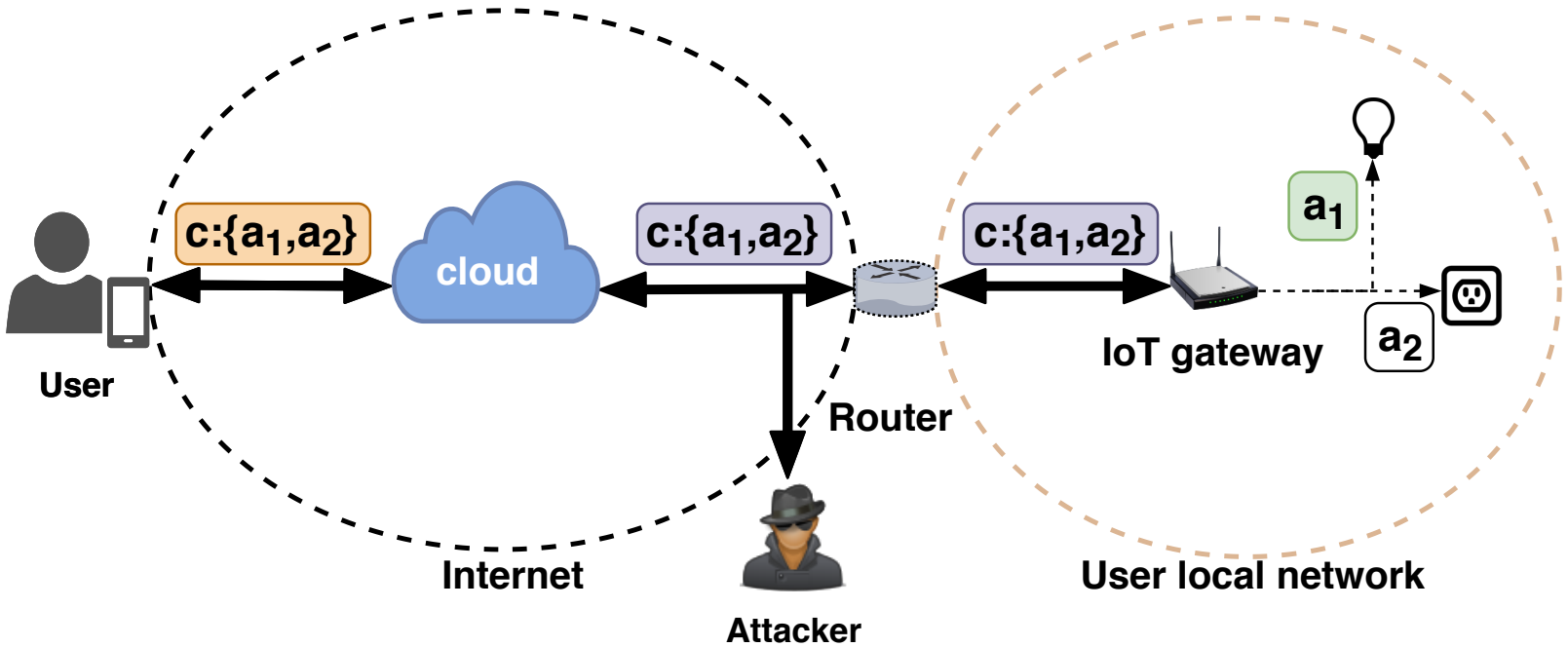
- Internet of Things (IoT) devices widely used
- Number of attacks increased by 600% between 2016 and 2017

  $\rightarrow$ Solutions to detect compromised IoT devices proposed
- IoT devices in smart homes $\rightarrow$ user privacy leakage

  $\rightarrow$ Close vicinity required

Introduction - IoT environment investigated

- command c = set of actions a_i



→ Measure the level of user privacy leakage exposed by an IoT gateway

Challenges and assumptions

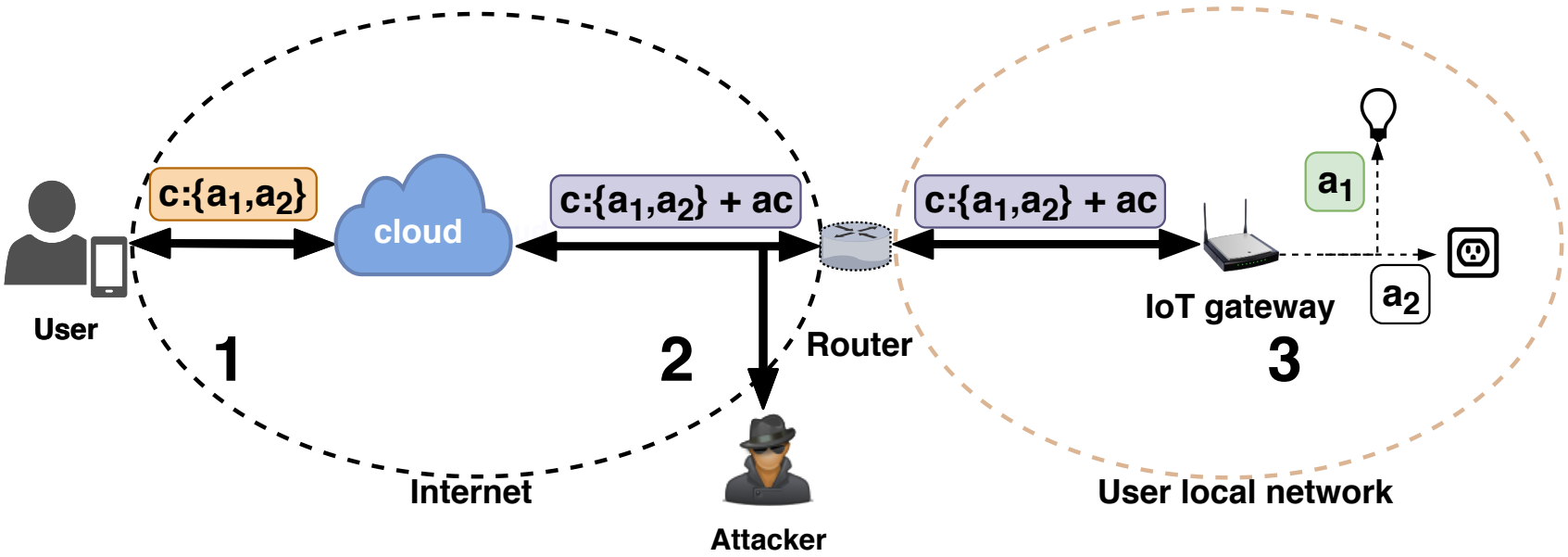
Our method raises some challenges:

• **Encryption** • **Gateway abstraction** • **Signature generation**

We made the following assumptions:

- **Actions data structure (1 action $\rightarrow$ 1 device)**
- **Command robustness**
- **Impact of the actions on the network packet sizes**
- **Similarities between user $\leftrightarrow$ WS and WS $\leftrightarrow$ IoT gateway**

Overview of our approach



1. From inputs user ↔ cloud (WS),
   extract features from WS ↔ IoT gateway network traffic
2. Learning of the signatures
3. User action identification

Conclusion and future work

- Gateway $\rightarrow$ security assessment harder
- Lack of privacy $\rightarrow$ actions performed, number of devices, device-type
- Collision $\rightarrow$ exact actions deduction harder

Future work:

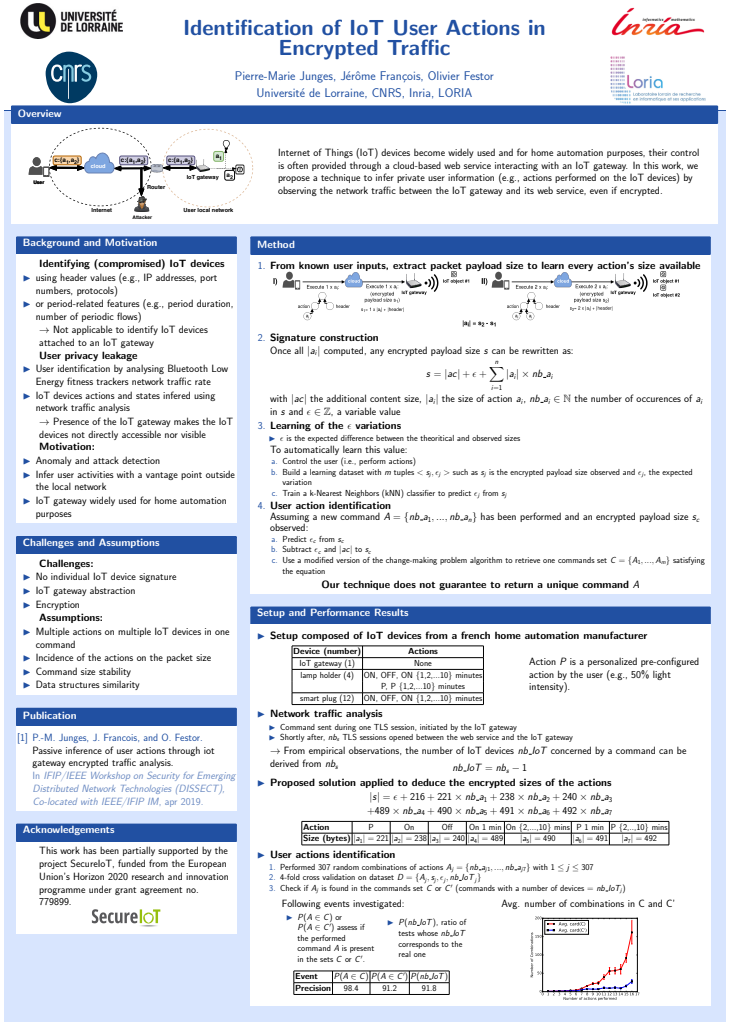
- Full automation of our method
- Apply our technique on other IoT gateways
- Create activity profiles to detect anomalies and attacks

Acknowledgments This work has been partially supported by the project SecureIoT, funded from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 779899.

SecureIoT

Thanks for listening !

Any questions?



Identification of IoT User Actions in Encrypted Traffic

Pierre-Marie Junges, Jérôme François, Olivier Festor
Université de Lorraine, CNRS, Inria, LORIA

Overview

Internet of Things (IoT) devices become widely used and for home automation purposes, their control is often provided through a cloud-based web service interacting with an IoT gateway. In this work, we propose a technique to infer private user information (e.g., actions performed on the IoT devices) by observing the network traffic between the IoT gateway and its web service, even if encrypted.

Background and Motivation

Identifying (compromised) IoT devices
- using header values (e.g., IP addresses, port numbers, protocols)
- or period-related features (e.g., period duration, number of periodic flows)

→ Not applicable to identify IoT devices attached to an IoT gateway

User privacy leakage
- User identification by analysing Bluetooth Low Energy fitness trackers network traffic rate
- IoT devices actions and states infered using network traffic analysis

→ Presence of the IoT gateway makes the IoT devices not directly accessible nor visible

Motivation:
- Anomaly and attack detection
- Infer user activities with a vantage point outside the local network
- IoT gateway widely used for home automation purposes

Challenges and Assumptions

Challenges:
- No individual IoT device signature
- IoT gateway abstraction
- Encryption

Assumptions:
- Multiple actions on multiple IoT devices in one command
- Incidence of the actions on the packet size
- Command size stability
- Data structures similarity

Publication

[1] P.-M. Junges, J. Francois, and O. Festor. Passive inference of user actions through iot gateway encrypted traffic analysis. In *IFIP/IEEE Workshop on Security for Emerging Distributed Network Technologies (DISSECT), Co-located with IEEE/IFIP IM*, apr 2019.

Acknowledgements

This work has been partially supported by the project SecureIoT, funded from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 779899.

Method

1. **From known user inputs, extract packet payload size to learn every action's size available**

$$|a_i| = s_2 - s_1$$

2. **Signature construction**
Once all $|a_i|$ computed, any encrypted payload size s can be rewritten as:

$$s = |ac| + \epsilon + \sum_{i=1}^{n} |a_i| \times nb\_a_i$$

with $|ac|$ the additional content size, $|a_i|$ the size of action a_i, $nb\_a_i \in \mathbb{N}$ the number of occurences of a_i in s and $\epsilon \in \mathbb{Z}$, a variable value

3. **Learning of the ϵ variations**
- ϵ is the expected difference between the theoritical and observed sizes

To automatically learn this value:
a. Control the user (i.e., perform actions)
b. Build a learning dataset with m tuples $< s_i, \epsilon_i >$ such as s_i is the encrypted payload size observed and ϵ_i, the expected variation
c. Train a k-Nearest Neighbors (kNN) classifier to predict ϵ_j from s_j

4. **User action identification**
Assuming a new command $A = \{nb\_a_1, ..., nb\_a_n\}$ has been performed and an encrypted payload size s_c observed:
a. Predict ϵ_c from s_c
b. Subtract ϵ_c and $|ac|$ to s_c
c. Use a modified version of the change-making problem algorithm to retrieve one commands set $C = \{A_1, ..., A_m\}$ satisfying the equation

Our technique does not guarantee to return a unique command A

Setup and Performance Results

- **Setup composed of IoT devices from a french home automation manufacturer**

Device (number)	Actions
IoT gateway (1)	None
lamp holder (4)	ON, OFF, ON {1,2,...10} minutes P, P {1,2,...10} minutes
smart plug (12)	ON, OFF, ON {1,2,...10} minutes

Action P is a personalized pre-configured action by the user (e.g., 50% light intensity).

- **Network traffic analysis**
  - Command sent during one TLS session, initiated by the IoT gateway
  - Shortly after, nb_s TLS sessions opened between the web service and the IoT gateway

→ From empirical observations, the number of IoT devices nb_IoT concerned by a command can be derived from nb_s

$$nb\_IoT = nb_s - 1$$

- **Proposed solution applied to deduce the encrypted sizes of the actions**

$$|s| = \epsilon + 216 + 221 \times nb\_a_1 + 238 \times nb\_a_2 + 240 \times nb\_a_3$$
$$+489 \times nb\_a_4 + 490 \times nb\_a_5 + 491 \times nb\_a_6 + 492 \times nb\_a_7$$

Action	P	On	Off	On 1 min	On {2,...,10} mins	P 1 min	P {2,...,10} mins
Size (bytes)	$\|a_1\| = 221$	$\|a_2\| = 238$	$\|a_3\| = 240$	$\|a_4\| = 489$	$\|a_5\| = 490$	$\|a_6\| = 491$	$\|a_7\| = 492$

- **User actions identification**
1. Performed 307 random combinations of actions $A_j = \{nb\_a_1, ..., nb\_a_{17}\}$ with $1 \leq j \leq 307$
2. 4-fold cross validation on dataset $D = \{A_j, s_j, \epsilon_j, nb\_IoT_j\}$
3. Check if A_j is found in the commands set C or C' (commands with a number of devices $= nb\_IoT_j$)

Following events investigated:
- $P(A \in C)$ or $P(A \in C')$ assess if the performed command A is present in the sets C or C'.
- $P(nb\_IoT)$, ratio of tests whose nb_IoT corresponds to the real one

Event	$P(A \in C)$	$P(A \in C')$	$P(nb\_IoT)$
Precision	98.4	91.2	91.8

Avg. number of combinations in C and C'

Thanks for listening !

Any questions?