ARGUS: Context-Based Detection of Stealthy IoT Infiltration Attacks

Phillip Rieger, Marco Chilese, Reham Mohamed, Markus Miettinen, Hossein Fereidooni, and Ahmad-Reza Sadeghi, *Technical University of Darmstadt*

https://www.usenix.org/conference/usenixsecurity23/presentation/rieger

This artifact appendix is included in the Artifact Appendices to the Proceedings of the 32nd USENIX Security Symposium and appends to the paper of the same name that appears in the Proceedings of the 32nd USENIX Security Symposium.

August 9–11, 2023 • Anaheim, CA, USA

978-1-939133-37-3

Open access to the Artifact Appendices to the Proceedings of the 32nd USENIX Security Symposium is sponsored by USENIX.

USENIX'23 Artifact Appendix: ARGUS: Context-Based Detection of Stealthy IoT Infiltration Attacks

Phillip Rieger Marco Chilese Reham Mohamed

Markus Miettinen Hossein Fereidooni Ahmad-Reza Sadeghi

Technical University of Darmstadt

A Artifact Appendix

A.1 Abstract

The continuous growth and expansion of Internet of Things (IoT) application domains, device diversity, and connectivity is a well-established trend. IoT devices have been implemented to manage and monitor various functions in smart homes, buildings, cities, and factories. However, this surge in adoption has made IoT devices an appealing target for cyber attackers. ARGUS analysis the IoT devices' behavior to detect contextual attacks, e.g., opening a smart lock while the smart home residents are absent. The artifact provides the 5 real-world datasets that were collected as part of the research. Using the datasets provides a benchmark dataset for future works and eases the comparison to future approaches.

A.2 Description & Requirements

The dataset contains the benign status updates of various IoT devices deployed for several months in 5 different real homes. The corresponding folder of a dataset contains for each day one CSV file, each with the devices' status updates of the corresponding day. The CSV files contain three columns, the time stamp, the unique identifier of the device/sensor, and the new status. The column new_status contains the status that the IoT device reported at the given time. Therefore, depending on the device, the value can be boolean (on/off), nominal (e.g., for the weather sunny, cloudy, partlycloudy, etc.) or numerical. The timestamps are given in UTC. However, it should be noted that the local time zone was ECT. Table 1 shows the number of events per dataset. In total, the artifact consists of 2 599 292 events.

Based on these events, attacks can be easily simulated, e.g., by injecting an event that sets the status of "camera.status" to on, while the sensor "person.home" has the status "home.

Table 1: Number of events contained in each dataset

Dataset	#Events
Home1	2 599 292
Home2	362 744
Home3	16 952
Home4	311 111
Home5	459 670

To obtain the dataset, we gathered data from IoT devices across various smart home environments. These environments, named Home 1- Home 5, were made up of multiple sensors (such as temperature, humidity, and motion sensors) and actors (such as light bulbs and thermostats). To ensure the dataset was diverse and that the individual setups differ from each other, each home also included additional sensors and actors. For instance, Home 1 had a CO_2 sensor, while Home 4 and Home 5 had multiple smart thermostats. The devices were installed in different homes, ranging from single-person apartments to shared homes with four inhabitants. In total, ten male and female participants from different age groups, including teenagers, students, and adults up to around 49 years old, were involved in the experiments. The data was collected using the open-source smart-home control system, Home Assistant.

The smart-home setups include 3 homes with multiple rooms and multiple inhabitants (Home 3, Home 4, Home 5), a one-room apartment (Home 2), as well as, a single room in a shared apartment (Home 1). The experiments included ten different male and female participants (teenagers, students, and adults up to approximately 49 years). We made use of the deployed home automation platform (in our setup HomeAssistant) to automatically trigger events, e.g., turning off the camera when the user comes home, or to turn off the heating when the window is opened.

Table 2 shows for each setup in the dataset, a detailed list of the deployed IoT devices and measured values. The measured values cover different categories of contextual features: i) Sensors/devices that measure ambient or temporal features (e.g., temperature, humidity, and luminosity), ii) user features (e.g., user presence and sleep confidence), and event features (e.g., states of the light bulbs, doors or windows).

Table 2: Deployed devices in the collected real-world IoT dataset. The deployment of a sensor/actor is indicated by ●, while the absence is indicated by ○.

Device	Home 1	Home 2	Home 3	Home 4	Home 5
Automation - All lights off	○	○	○	○	●
Automation - All lights on	○	○	○	○	●
Automation - Camera off when at home	○	○	○	●	○
Automation - Dinner lights	○	○	○	●	●
Automation - Dinner table light	○	○	○	○	●
Automation - Gaming mode	○	○	○	○	●
Automation - Heating boost off	○	○	○	○	●
Automation - Light off when no motion	○	○	○	●	●
Automation - Lights off in the evening	○	●	○	○	○
Automation - Lights off when too bright	○	●	○	●	○
Automation - Lights on in the morning	○	●	○	○	○
Automation - Lights on when motion detected	○	●	○	●	○
Automation - Piano Light	○	○	○	○	●
Automation - Sofa Lamp	○	○	○	○	●
Automation - Studio Light off	○	○	○	○	●
Automation - Studio Light on when motion	○	○	○	○	●
Automation: Camera on when user leave	○	○	○	●	○
Camera Status Sensor	○	●	●	○	○
Climate - Control access point 1	○	○	○	○	●
CO_2 Sensor Status	●	○	○	○	○
CO_2 Sensor	●	○	○	○	○
Control Access Room 1 Sensor	○	○	○	●	○
Door Sensor	●	●	●	●	●
Floor lamp	○	○	○	○	●
Heating - heater valve	○	○	○	○	●
Heating Temperature Sensor	●	●	○	●	○
Homematic - Radiator Thermostat Temperature Sensor	○	○	●	○	○
Humidity Sensor	●	●	○	●	●
IKEA Tradfri Roller Blind Sensor	●	○	○	○	○
IP Camera - Light Level	○	○	○	○	●
IP Camera - Motion	○	○	○	○	●
IP Camera - Motion Active	○	○	○	○	●
IP Camera - Pressure	○	○	○	○	●
IP Camera - Sound	○	○	○	○	●
Lamp consumption	○	○	○	○	●
Lamp consumption (daily)	○	○	○	○	●
Lamp consumption (total)	○	○	○	○	●
Lamp current	○	○	○	○	●
Lamp voltage	○	○	○	○	●
Light - Ceiling	●	●	●	●	●
Light - Desk Lamp	●	●	●	●	●
Light - Living Room	○	○	●	○	○
Philips Hue - Light Level Sensor 1	○	●	●	●	●
Philips Hue - Light Level Sensor 2	○	○	●	○	○
Philips Hue - Motion Sensor 2	○	○	●	○	○
Philips Hue - Temperature Sensor 1	○	●	●	●	●
Philips Hue - Temperature Sensor 2	○	○	●	○	○
Philips Hue - White Lamp 2	○	○	●	○	○
Philips Hue - White Lamp 3	○	○	●	○	○
Philips Hue - Motion Sensor 1	○	○	●	●	●
Piano lamp	○	○	○	○	●
Radiator Thermostat Sensor	○	○	○	●	●
Smartphone - Battery Life	○	●	○	○	○
Smartphone - Charging	○	○	○	○	●
Smartphone - Charging Sensor	●	○	○	○	○
Smartphone - Connected to WLAN	○	●	○	●	○
Smartphone - Detected Activity	●	●	○	○	●
Smartphone - Light Sensor	○	●	○	○	○
Smartphone - Locked	○	○	○	○	●
Smartphone - Phone Status	○	●	○	○	○
Smartphone - Sleep Confidence	●	●	○	●	○
Smartphone - Sleep Segment	○	○	○	●	○
Smartphone - Tracker	●	●	○	●	○
Studio lamp	○	○	○	○	●
Sun Sensor	●	●	●	●	●
Temperature Sensor (ESP)	●	●	○	●	●
User Presence	●	●	○	●	○
Weather - Home Location	●	●	●	●	●
Weather - Town	●	○	●	●	○
Window Sensor	●	●	●	●	○

A.2.1 Security, privacy, and ethical concerns

The dataset collection raised ethical concerns, as the recorded behavior of the users might contain sensitive data. We addressed these concerns by ensuring that all affected persons, i.e., the users as well as all guests, were aware of the data collection and gave their consent. Further, we limited the approach to non-privacy-sensitive sensors and excluded the other sensors like the geolocation or the SSID of the WiFi network that the mobile phone is connected to. In addition, all potentially sensitive data items were anonymized. Our experimental set-up has been reviewed and approved by the ethics board of our university.

As the artifact is an anonymized dataset, it does not raise any security, privacy, or ethical concerns for people using this dataset. The license allows the usage of any non-commercial purpose (CC-BY-NC-SA).

A.2.2 How to access

The dataset was uploaded at GitHub and is accessible at: https://github.com/TRUST-TUDa/argus-data/tree/606d5a5ebe78f602e27b9f2c48ea103348463eeb. The dataset that was used for the paper is tagged as ArtifactAppendix.

A.2.3 Software dependencies

The git program to check out the dataset from GitHub.

A.3 Set-up

The artifact consists of the datasets that were collected for the paper. The artifact includes contextual events from several real-world smart homes and makes them publicly available for future research. Therefore, the artifact does not include any software/code but is intended as a benchmark for future work on contextual intrusion detection.

A.3.1 Installation

Clone the dataset from GitHub and checkout the specified state via:

```
$ git clone https://github.com/TRUST-TUDa/argus-data.git
$ cd argus-data
$ git checkout 606d5a5
```

A.3.2 Basic Test

Verify via `git status` that the repository is at 606d5a5 to verify that all files were cloned correctly.

A.4 Version

Based on the LaTeX template for Artifact Evaluation V20220926. Submission, reviewing and badging methodology followed for the evaluation of this artifact can be found at https://secartifacts.github.io/usenixsec2023/.