itations are threefold. First, the logic of some smart apps is

Smart Apps

SmartThings

too complex to be mined accurately, causing false negatives

Events

Commands

Cloud

Device Handler

and positives. For example, the event pattern introduced by

the smart app logic "Turn off a smart plug 30 minutes after

Network

two motion sensors in the living room are both motionless" is

Zigbee

ZWave

WiFi

Connection

difficult to be mined considering the 'AND' logic between

two motion sensors and the 30 minutes action delay. As a re-

IoT Device

sult, an anomaly "the smart plug fails to turn off" may not be

Microcontroller

Wireless Module

Cyber Part

detected. Second, the learning results are typically difficult

IoT Device

to interpret; thus, they can hardly be explained and often

Physical Part

confuse users. Third, the learning results cannot be updated

Thermometer Light Bulb

Relay

quickly when smart apps or configuration changes. A long

Figure 2: The SmartThings architecture.

re-training process is then needed to adapt to the changes

and many false alarms arise before the re-training is done.

and can be refined easily to resolve conflicts with smart

Intuitively, incorporating semantic information, such as

apps and updated conveniently when apps change.

automation logic, device types, relations and installation lo-

We propose the notion of shadow execution for smart

cations, can help improve the accuracy of anomaly detection.

homes, which simulates the normal behaviors of a home

However, there are a number of challenges to overcome in

according to the learned correlations and detects anoma-

order to realize this idea: 1) Standard data mining methods

lies at a fine granularity, i.e., IoT events.

take event logs as inputs; however, it is unknown how to

represent the diverse semantic information in the form of

We implement a prototype HAWatcher and evaluate it

event logs. 2) System behavior patterns derived from smart

on four real-world testbeds. HAWatcher reaches a high

apps and those mined from events logs may conflict. It is

precision of 97.83% and a recall of 94.12%, significantly challenging to identify and resolve these conflicts. 3) When

outperforming prior approaches.

smart apps change, there are no effective methods to update

the system profiling accordingly.

The rest of the paper is organized as follows. In Section 2,

To fill the gap, we present Home Automation Watcher

we describe background about appified smart homes. In Sec-

(HAWatcher), a novel anomaly detection system for appified

tion 3, we survey IoT device anomalies and present the threat

home automation systems. We propose a semantics-assisted

model. In Section 4, we describe three correlation channels

mining method that exploits diverse semantic information

and the representation of correlations. We present the design

to construct hypothetical correlations (where a correlation de-

details in Section 5. The evaluation is presented in Section 6.

scribes how a device state or event correlates with another),

We discuss related work in Section 7, and limitations and

and use event logs as evidence to verify them. Second, as

future work in Section 8. The paper is concluded in Section 9.

the correlations are explainable according to the semantics,

2

Background: Appified Smart Homes

they can be easily refined to resolve conflicts with smart

apps. Third, still thanks to explainability, they can be up-

IoT devices in smart homes have become increasingly inte-

dated conveniently according to smart app changes. The grated via IoT platforms for rich automation. IoT integration

correlations are then used by our shadow execution module

platforms, such as SmartThings, Amazon Alexa, and Open-

to simulate normal behaviors in the virtual world. The simu-

HAB, support trigger-action automation programs. On these

lated states are compared to those in the real world through

platforms, despite the huge number of IoT devices, they are

both contextual checking and consequential checking, and

abstracted into a small number of abstract devices. For ex-

inconsistencies during comparison are reported as anomalies.

ample, a smart light, regardless of its brand, shape, size, and

We make the following contributions.

wireless technology, is abstracted into the same abstract de-

vice, light. Each abstract device has its associated events and

We propose a novel anomaly detection solution for appi-

commands. Device vendors can have their products support

fied smart homes. It meets the emerging need of detect-

integration by realizing the events and commands.

ing anomalies caused by IoT malfunctions or attacks.

We choose SmartThings [21] as an example IoT integra-

tion platform to present our design, as SmartThings is one of

We propose a semantics-assisted mining method, which

the leading platforms and supports sophisticated automation

infuses various semantic information (smart apps, con-

logic. Other integration platforms, such as Amazon Alexa,

figuration, device types, installation locations) into the have similar structures. As illustrated in Figure 2, a typical

mining process. An NLP-based approach is developed

SmartThings deployment has a cloud-centric architecture of

to describe device relations for generating hypothetical

four layers. On the top is the SmartThings cloud, where smart

correlations. The mined correlations are explainable,

apps run and interact with abstracted capabilities. The cloud