Characterizing and Classifying IoT Traffic in Smart Cities and Campuses

Paper's analysis
ITPA 2019-2020

Andrea Graziani - 0273395

Università degli Studi di Roma "Tor Vergata"
FACOLTA' DI INGEGNERIA
Corso di Laurea Magistrale in Ingegneria Informatica

November 30, 2020

Research's goal

According to Sivanathan et al. [1], research's goal is to:

> *"[...] develop a classification method that can not only distinguish IoT from non-IoT traffic, but also identify specific IoT devices with over 95% accuracy."*

What is the reason according to which is important to profile IoT traffic?

1. To understand IoT devices "*normal*" **traffic pattern** in terms of their **activity pattern** (traffic rate, idle durations, etc.) and **signalling overheads** (`DNS`, `NTP`, etc.).
2. To enhance **cyber-security** involving IoT devices which administration belong to **different authorities**.
   According to Sivanathan et al. [1], is possible to improve security deploying a network-level security mechanisms which, analysing traffic patterns, is capable to **identify attacks knowing the normal traffic pattern of monitored IoT devices**.

Research's goal

In other words, research's goal is to build an classification model for IoT devices based on **machine learning** techniques, which building passes through following steps:

1. Collect data from an IoT environment.
2. Characterize traffic pattern corresponding to the various IoT devices.
3. Develop a classification technique that learns the behaviour of an IoT device and is able to identify it based on its traffic pattern.

We believe that the "Smart environment" by researchers is not suitable for to build an IoT traffic trace in order to classifiyn IoT traffic in smart cities and campuses for security purposes.
We are convinced that model not
Due to the **type** and the **number** (only 20 devices!) of IoT devices used for their experiments, adopted "Smart environment" is more suitable for a smart home rather than a smart city or campus as stated by authors.

Data-set building

- Since Sivanathan et al. [1] adopted a **supervised machine learning algorithms** to build their model, is necessary to generate a **data-set**, in order to provide an appropriate input for the **learning phase**.

To be precise, Sivanathan et al. [1] built a **time series** dataset, where each instance is indexed by time.
Each data-set's instance contain following **attributes** (or **features**):

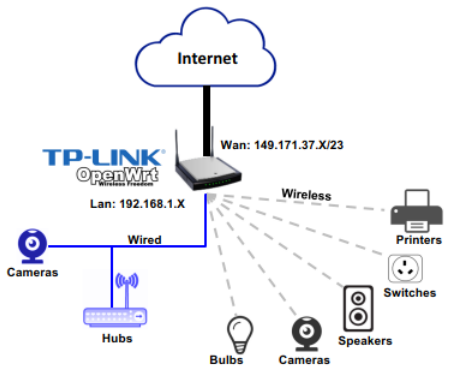
- Traffic load.
- Average packet size
- Protocol used and port number
- active and sleep times
- unique DNS requests,
- NTP interval,
- peak/mean rate

"Smart environment"'s analysis: Overview

In order to building needed dataset, simulating a real usage scenario, Sivanathan et al. [1] built a so-called "*smart environment*", including:

- 21 unique IoT devices representing different categories, like **cameras**, **healthcare devices**, **hub**, **air quality sensors** and so on.
- A router, the TP Link Archer C7[1]
- Several non-IoT devices were also used, such as laptops, mobile phones and tablet.

Figure: "Smart environment"'s scheme



[1] https://www.tp-link.com/it/home-networking/wifi-router/archer-c7/#overview

"Smart environment"'s analysis: Observations

Observation # 1

The experimental smart environment build by researchers is characterized by a **star network topology**.

This is a very important observation, because a star topology allows us to:

- **Preserve battery life** of IoT devices because they **do not have to forward** other nodes data; in other words, *any IoT device receives, or transmits, only its own data*.
- **Decrease the complexity** of the network.

The LPWAN example

The implementation of `LoRaWAN` network is based on the star network topology, and mostly, stars-of-stars network.
As known `LoRaWAN` network belongs to `LPWAN` category, *which are specifically designed to achieve the need for low power, long-range, low bit error rate, and low cost* needed in IoT context.

"Smart environment"'s analysis: Observations

Observation # 2

Researchers use `IEEE 802.11` as **media access control** (`MAC`) and **physical layer** (`PHY`) protocol.

Observation # 3

Researchers did *not* specify which **version** of IEEE 802.11 standard has been effectively used.
We don't know with which **frequencies** data has been transmitted.

According to vendor's specifications regarding the TP Link Archer C7, is possible to know that the aforementioned router supports following protocols:

- `IEEE 802.11ac/n/a` at 5 GHz
- `IEEE 802.11n/b/g` at 2.4 GHz

"Smart environment"'s analysis: Observations

Observation # 4

We believe that above protocols are **not** fully optimized for IoT business models and devices used in smart cities and campuses for following reasons:

- These technologies provide a **short/medium coverage** with **100-to-1000** meters range. Provided coverage range can be not enough to fulfil all use cases.
  - This is due to **mid/high frequencies** used by these protocol which are **vulnerable to several side effect during signal propagation** (*blocking*, *reflection*, *refraction* and so on)
- As stated by Sivanathan et al. [1] too, IoT devices require **low power** and less **data-rate**.
- They are affected by overhead caused by short packets transmission which are very common is many IoT scenarios.

"Smart environment"'s analysis: Observations

Utilizing sub-1 GHz bands used `802.11ah` or `LoRaWAN` by to provide better propagation characteristics in outdoor scenarios. Low frequencies signal are less affected by obstacles presence.

A combination of low-band, mid-band and high-band spectrum is desirable to manage all possible use cases.

	802.11ac	**802.11n**	**802.11a**	**802.11ah**	**LoRaWAN**
Frequency (GHz)	5	2.4,5	5	0.7/0.8/0.9	$0.863 - 8.70(EU)$
Sensitivity (dbm)	-82	-82	-88	-98	$[-124, -137]$
Bit rate (Mb/s)	6.5	6.5	1.5	0.15	$[293, 5469]$
Max coverage range (km)				$\backsim 1$	$\backsim 15$

IoT Traffic

According to their experimental results, Sivanathan et al. [1] stated that:

> *"[...] if we consider only the load imposed by the IoT devices, then there is a dramatic reduction in the peak load (1 Mbps) and average loads (66 Kbps), [...], implying that traffic generated by IoT devices is small compared to traditional non-IoT traffic."[1, par. IV.A]*

> *"the traffic pattern of one IoT device [...] a pattern of active/sleep communication emerges. [...] IoT active time [...] decays rapidly initially (only 5% of sessions last longer than 5 seconds), with the maximum active time being 250 seconds in our trace. This shows that IoT activities are short-lived in general."[1, par. IV.A]*

Observation # 1

Several IoT devices used by Sivanathan et al. [1] for their experiments **are battery operated**.
For instance:

- The *Withings Smart scale* device is powered by 4 1.5 V alkaline cells (AAA).[a]
- Similarly, the *Netatmo Weather station* device is powered by 2 1.5 V alkaline cells (AAA) with an **estimated autonomy of about 2 years**.[b]
- The *Blipcare blood pressure meter* device is powered by an internal battery.[c]

---

[a] `https://www.withings.com/it/en/body`
[b] `https://www.netatmo.com/it-it/weather/weatherstation/specifications`
[c] `http://www.blipcare.com/`

Since many IoT devices are battery powered, **maximize energy efficiency**, in order to **preserve devices lifetime**, is critical.

As expected, observing the figure reported below, is very easy to understand that the power management approach, used to preserve battery life, is based on **periodic sleep**, during which radio transceiver are turned off.

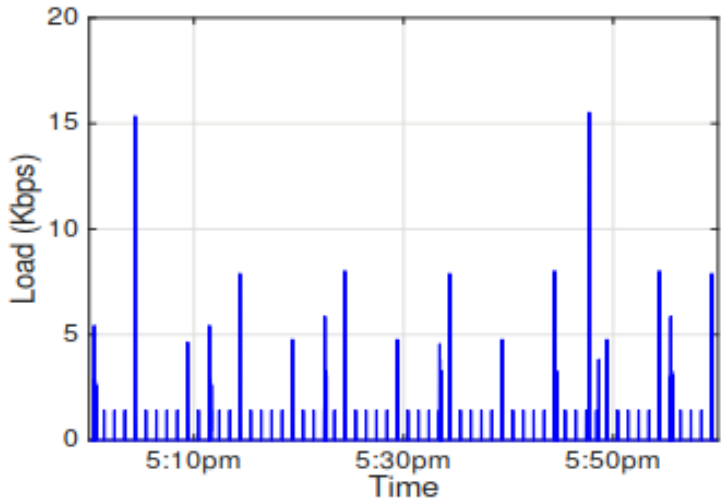


Figure: Load of LiFX light bulb device.

- According to Sivanathan et al. [1]'s report,

> *"Nevertheless, about 45% of IoT traffic (by number of packets) is not sent over HTTPS to the servers on the public Internet indicating that a sizeable fraction of IoT traffic is not being securely transported over the Internet."[1, par. IV.B]*

Another challenge facing IoT successful deployment is the lack of a universal platform, protocol, and a programming language. Today, all connected devices follow a different set of protocols and platforms. The need of the time is to have collaboration among the connecting devices. For this, large enterprises such as LG, Samsung and Philips etc. should join hands and make a consortium for the development of universal coding language and platform. The proposed solution can solve the compatibility issues of IoT up to a significant extent

Another challenge facing IoT successful deployment is the lack of a universal platform, protocol, and a programming language. Today, all connected devices follow a different set of protocols and platforms. The need of the time is to have collaboration among the connecting devices. For this, large enterprises such as LG, Samsung and Philips etc. should join hands and make a consortium for the development of universal coding language and platform. The proposed solution can solve the compatibility issues of IoT up to a significant extent

- Why `HTTPS` and `HTTP` are the dominant protocols used by IoT devices, as stated by Sivanathan et al. [1]?
  - As stated by **WOT**, HTTP/HTTPS protocols play a very important role into **Web of Things Architecture** for several reasons:
    1. Facilitate both the **integration of IoT devices** with existing services currently available on the Web and the **Web applications development** exploiting **REST architectural style**
    2. They offer a **direct access** for users to IoT devices data and services, without the need for installing additional software.
       In fact, using a Web browser (or any HTTP library in the case of a software client) client are able to to directly extract, save and share smart things data and services.
       This ensures the usability of the architecture and minimizes the entry barriers for final users.

- According to Sivanathan et al. [1], the set of IoT devices used for their experiments including a huge amount of **sensors**, including air quality sensors and health-care devices.
  As known, aforementioned kind of devices generate a huge amount of data modelled as **time series**, that is an array of values indexed by time.
  According to **TIMESERIES**, the stream of data generated by all these IoT sensors is generally interfaced with database, through a so-called *southbound* interface, using HTTP RESTful protocol.
  Similarly, all applications requiring access to the data stored in the database, using the same protocol, through a so-called northbound interface.

Some references

[1] A. Sivanathan et al. "Characterizing and classifying IoT traffic in smart cities and campuses". In: *2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. 2017, pp. 559–564. DOI: 10.1109/INFCOMW.2017.8116438.