Mehran University Research Journal of Engineering and Technology
Vol. 40, No. 2, 335 - 345, April 2021
p-ISSN: 0254-7821, e-ISSN: 2413-7219
DOI: 10.22581/muet1982.2102.08

Clustering of IoT Devices Using Device Profiling and Behavioral Analysis to Build Efficient Network Policies

Muhammad Hamza[1a], Syed Mashhad M. Geelani[1b], Qamar Nawaz[2], Asif Kabir[3], Isma Hamid[4]

RECEIVED ON 25.08.2019, ACCEPTED ON 15.09.2020

ABSTRACT

The Internet of Things (IoT) has emerged as a new paradigm, and billions of devices are connected with the internet. IoT is being penetrated in major domains of daily life like health care, agriculture, industry, smart homes and monitoring of the environment. The operator of such complex, huge and diverse heterogeneous networks may not even be fully aware of their IoT devices working, activity, behavior and resource utilization *etc*. The efficient management of IoT devices becomes a challenge for network managers to ensure smooth network operation. Network traffic analysis of IoT devices is a necessary and rudimentary tool to understand the behavior of devices. In this paper firstly, we identify insights of device network traffic, discuss the activity patterns of some IoT devices and present a visual description of the pattern of IoT devices. Secondly, after analyzing the device's behavior, we build and demonstrate a profile of each device based on its activity cycle and traffic patterns information. Thirdly, the K-Means clustering algorithm is used to make clusters of IoT devices using their profile information. The clustering algorithm groups similar devices in a single group. The obtained results clearly describe the patterns of devices which help the network managers to make appropriate network policies for efficient secure network management.

Keywords: Internet of Things, Network Traffic Analysis, IoT Devices Clustering, Device Profiling, IoT Traffic Characteristics.

1. INTRODUCTION

Internet of Things (IoT) is commonly used as an umbrella term, in which internet and web is connected to a physical domain to develop smart environment [1]. IoT is also known as the industrial internet [2]. IoT penetrates in almost every domain. Society is shifting towards the "always connected" model [3]. The basic idea of IoT is to connect all the things in the world with the internet. Most of the things (devices) are supposed to be intelligent thus they are called smart objects [4], which are capable to identify, sense event, interact with other devices and make a decision themselves [5]. According to the statistical reports published by International Data Corporation (IDC), worldwide IoT devices data will grow 33 Zeta Bytes (ZB) to 175 ZB from 2018 to 2025 [6]. Another report [7] on IoT devices published by the Information Handling Services (HIS) anticipates that the number of devices would reach ~75 Billion by 2025, as shown in Fig. 1. In 2020 number of IoT devices connected with internet reach ~50 Billion [8]. IoT combines current Internet infrastructure with the latest technologies to ensure smooth, seamless efficient

[1] University Institute of Information Technology, PMAS- Arid Agriculture University, Rawalpindi, Pakistan. Email: [a]hamzachaudary20@gmail.com, [b]mushhad@uaar.edu.pk (Corresponding Author)
[2] Department of Computer Science, The University of Agriculture, Faisalabad, Pakistan. Email: qamar@uaf.edu.pk
[3] Department of Computer Science and Information Technology, University of Kotli, AJK, Pakistan. Email: asifkabir@cqu.edu.cn
[4] Department of Computer Science, National Textile University, Faisalabad, Pakistan. Email: ismahamid@ntu.edu.pk

interconnection between hundreds of billions of embedded systems [9].

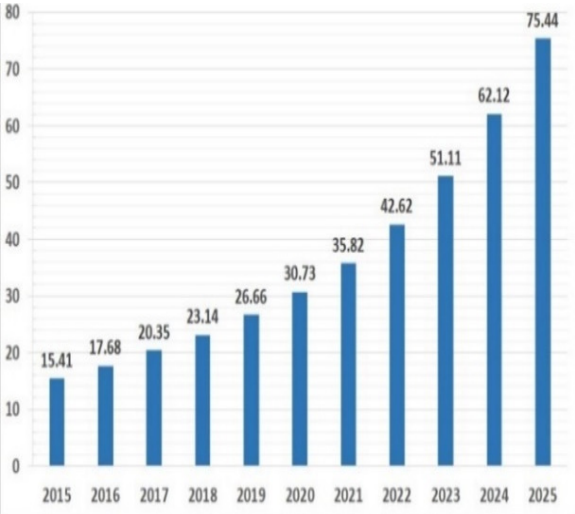

Fig. 1**:** Forecasting of IoT Devices Growth
Source: IHS.

Labeling of massive network traffic is a very tough task. To overcome such a problematic job the clustering is used in recent research to make clusters for IoT devices traffic [5]. Clustering has emerged as a vibrant technology for self-directed, smart and intelligent network administration and operation [10]. Clustering is widely used in multi-domains for data mining, finding hidden patterns of unlabeled diverse data [5].

IoT systems generate massive data which causes internet traffic to become more complex and heterogeneous. Internet traffic classification is considered as the most fundamental function of modern network management [11]. The huge traffic flow of IoT network traffic arises many challenges like device behavior, device profiling, network privacy [12], scaling, Manufacturer Usage Description (MUD) and standardization of IoT devices [13]. The IoT communication has changed the pattern of network traffic and increased the ratio of smart devices traffic on the internet. While estimating the huge number of connected things researchers do not even have a methodology to even count the exact number of devices and types of things [14]. In such an environment, it is difficult to identify and analyze the diverse behavior of IoT devices traffic from the network. Most IoT devices send constant or periodic traffic and make predictable patterns and devices to repeat these patterns regularly throughout traffic transmission. IoT devices require a limited number of protocols for their applications with a limited number of ports for transmission. Another issue of IoT devices is a closed communication pattern with pre-defined parameters set by their manufacturers while non-IoT devices proliferate hundreds of Domain Name System (DNS) packets to different number of domain servers openly. There is a need to incorporate data mining approaches to analyze the patterns of IoT generated traffic in order to resolve the above mentioned challenges.

In this paper, we propose a model that is able to classify the traffic of the devices by observing traffic packets and identify whether they are generated by the IoT or non-IoT device. Firstly, the proposed model examines network traces patterns of each IoT device and demonstrates some traffic patterns in a graphical form. Then a profile of IoT device is made on the basis of traffic pattern generated by the IoT device. In the end, clustering of IoT devices is done by employing device profiling. This will explore hidden common patterns of IoT devices and helps to make effective network polices to enhance the insight visibility of IoT based network traffic.

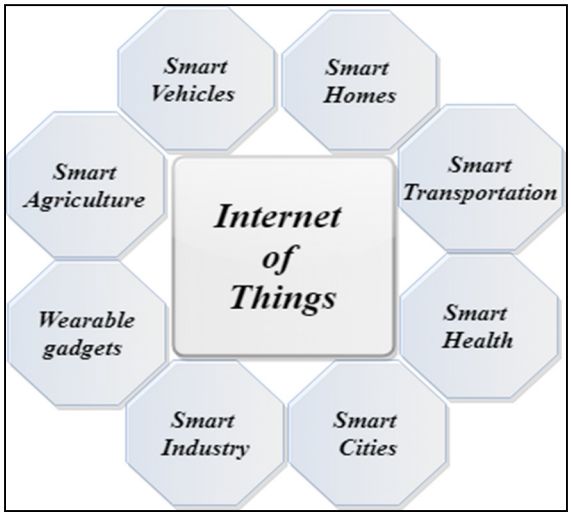

Fig. 2: IoT Applications

The Internet becomes a ubiquitous part of modern society. IoT is expanding globally and providing several benefits in almost every aspect of life as shown in Fig. 2. In the past only rooms and offices were

smart, now homes and campuses are becoming intelligent [15]. The scope of IoT is very huge, encompassing intelligent transportation, medical, surveillance, shopping and environment monitoring [16]. The industry is now equipped with hundreds of IoT devices to make it smarter. Moreover, many countries are planning to make their cities smart [17].

Clustering is considered one of the most common Machine Learning techniques. Clustering learns from unlabeled data because it is relatively easier than other data mining approaches, it divides unlabeled data into groups on the basis of some similarity [18]. The most recent researchers focus on IoT device-generated traffic fingerprinting due to the rapid increase of IoT devices traffic. In [19], researchers analyzed the device's traffic, discussed some insights of data that they found in the traffic. Their results clearly showed that traffic generated by devices was somehow different from each other. Moreover, they applied six different supervised algorithms on the traffic dataset and evaluated the results. Overall, Random forest gave high accuracy results. However, they used a small number of devices *i.e.* 4 that is unrealistic for a real network.

Recently, the necessity of traffic classification, application identification and device categorization has attracted various research efforts. Based on the statistical feature analysis, new research employs Machine Learning (ML) algorithms for making efficient traffic classifiers. The ML-based traffic classification achieves high accuracy and becomes a prominent structure [20]. In a smart environment, some devices fall in the IoT category and they produce heterogeneous data with various sizes of volume and format in which network traffic is varied [5]. The amount of data transfer by some IoT sensors is very less and does not require a high-speed link, devices like smart meters are expected to generate 0.07 MB per day [21]. In [22], a hybrid approach was applied, it focused on features extracted from the Transmission Control Protocol (TCP) sessions and used Random Forest (RF) classification and K-means for attribute clustering and finally characterized the IoT traffic.

Recently [23] proposed a cascade model to automatically identify the semantic type of devices using neural networks. It classifies devices in four classes with less than 75% accuracy. Researchers in [4] collect and synthesize network traffic traces from a smart environment and categorize 20 IoT devices in four classes with 95% accuracy. However, it did not provide the behavior analysis of IoT devices. The authors of the paper [3], describe an approach of ML for device identification. Initially, feature extraction was performed on TCP sessions (from SYN to FIN). The classifier used the metadata of packet and payload information for classification. They achieved an accuracy of 99.281% for identifying the IoT device. While classifier waited for the end of the individual device session to classify the device into a specific class. They had a very small dataset that contained only a few devices. Although they contributed to this field, problems are big because thousands of new diverse devices are connected on the internet daily.

In the article [7], the authors describe that it is possible to identify specific devices with network traffic traces even if the device enables protection methods like randomization of Media Access Control (MAC) addresses. This can be achieved with one feature protocol looking at DNS requests to recognize the type of devices domain that is requesting. Moreover, many of the traffic patterns mentioned in this paper hold true in our research. Devices flow statistics are demonstrated in this research which shows the behavior of IoT devices but only with reference to bandwidth. The contribution of the researcher is very helpful for future work in the IoT network domain, however, they focus only on the bandwidth aspect [24]. Brief comparison of our proposed work and some recent related work are described in Table 1.

The recently published material concludes that there is still room for expansion for device profiling and categorization. Some aspects are shown in Fig. 3. Some techniques need prior knowledge of data for classification which needs labeling. They first make MUD profile of each device then monitor IoT devices activities and verify device behavior in network [25]. However, data labeling is a very tough task. IoT device traffic increases the volume and complexity of the network. To overcome the current issues there is a need for an analysis of IoT device's behavior.

2. MATERIALS AND METHODS

Our proposed approach uses common attributes which are easily extracted from network traffic traces, even traffic is encrypted. To start the experiment, we first identify the traffic patterns, data types, and data rate caused by different IoT devices. For this purpose, a smart environment equipped with different IoT devices is required to capture the traffic traces of devices. The detailed methodology is depicted in Fig. 4. The data was extracted passively, the smart campus is developed for research purposes in [4]. The real data was captured continuously for 14 days , which is available in [26]. The size of network traffic data

Table 1: Comparison between Related Work and Proposed Work

Paper Reference	**Year**	**Methodology / Attributes**			**Advantages / Parameters**				
		Proposed Approach	# of Devices	Flow Info (F) Header Info (H)	Identify Device	Device Behavior Analysis	Device Profiling	Categorize Devices	IoT Traffic Characteristics
Proposed Work	2019	Script Clustering	20	F/H	Performed	Limited	Average	Cluster	Only Tested Characteristics
Hamza, A *et .al.* [25]	2019	Script	28	F/H	Performed	Yes	Yes	N/A	No
Lei B *et. al.* [23]	2018	Classification LSTM-CNN	16	F/H	Performed	No	No	Classify	No
Hamza, A *et.al.* [13]	2018	Script	28	H	Performed	Yes	Basic	N/A	N/A
Y.Amar *et.al.* [24]	2018	Script	14	F/H	No	Limited	Essential	N/A	No
M.R *et. al.* [19]	2018	Classification	4	F	Performed	No	No	Classify	No
Arunan, S. *et. al.* [21]	2017	Classification Clustering	20	H	Performed	No	No	Classify	Few Basics
Y. Meidan *et. al.* [1]	2017	Classification Script	17	F	Performed	No	No	Classify	No
Miettinen *et. al.* [12]	2017	Classification Script	27	H	Performed	Limited	No	Classify	No

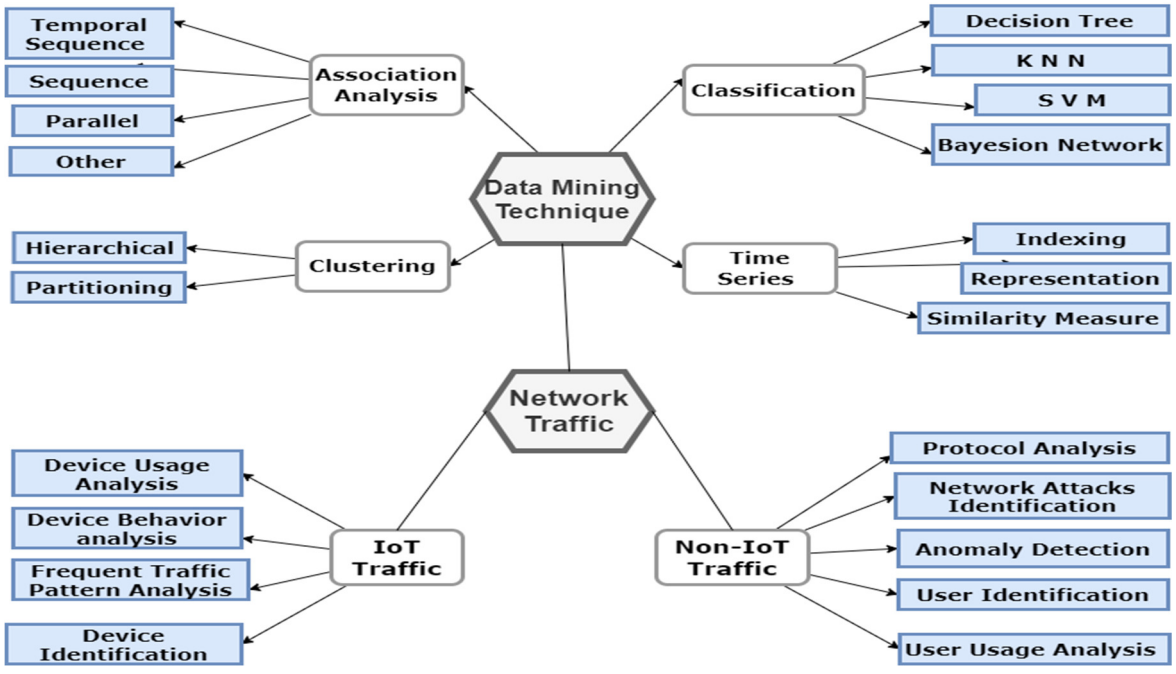

Fig. 3: Taxonomy of IoT Data Mining

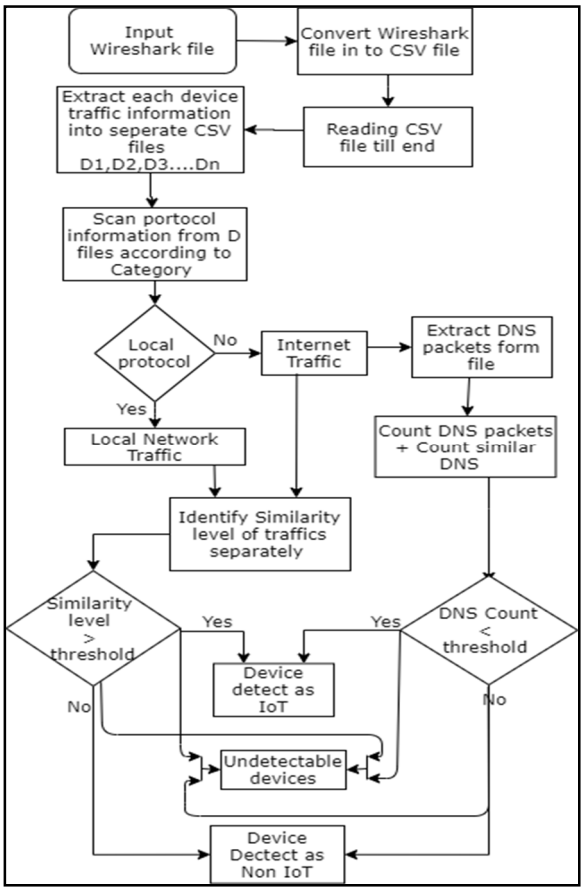

Fig. 4: Proposed Methodology

was 8 GB which was enough to get reasonable results. There are 20 devices connected in the network including both IoT as well as Non-IoT devices as shown in Fig. 5.

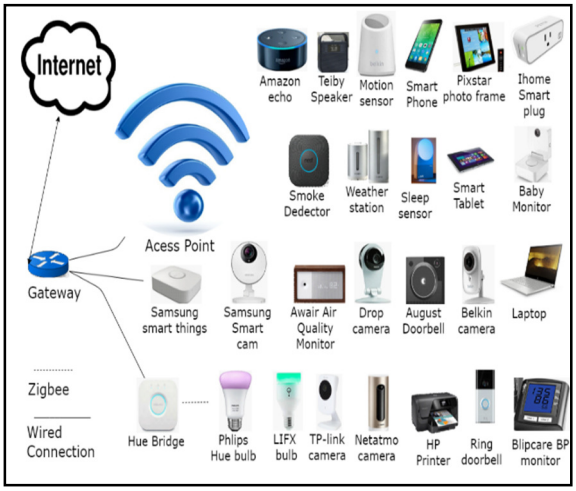

Fig. 5: Experimental setup for IoT devices

Table 1: Tools and Technology

Tools	Purpose	Reference site
Wireshark	Network Traffic capturing	www.wireshark.org/.
Python	Script Coding	www.python.org/
Origin	Graphical Representation	https://www.originlab.com
Weka	Data Mining Clustering	www.cs.waikato.ac.nz/ml/weka/.

Network traffic traces are shown in Fig. 6. Network traffic packets are captured with the help of the Wireshark tool listed in Table 2. The features selected

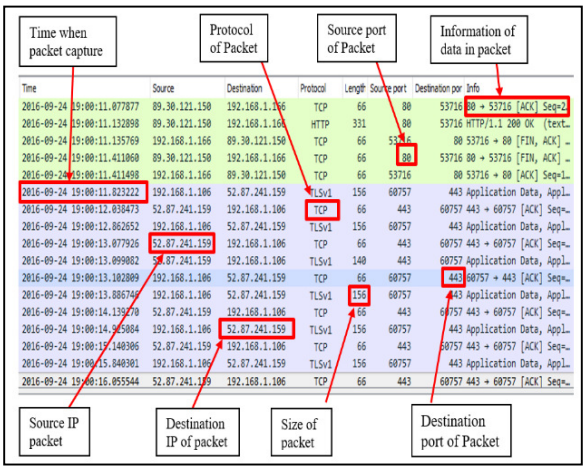

Fig. 6: Network Traffic Traces in Wireshark

from the flow are organized in the form of tuples comprised of time, packet size, protocol, source MAC, destination MAC, source port, destination port source IP, destination IP and information carried by packets. Some packets are shown in Fig. 6. All traffic packets are transformed into the dataset (.csv files). Other useful information was extracted from the dataset using Python and Weka listed in Table 2 to calculate the volume of packets, standard deviation, minimum and maximum packet size, average packet size and inter-packet, *etc*. These values are helpful for further data processing and statistical analysis.

With the help of a script applied on the dataset we split traffic traces on a base of MAC address. Once all devices traffic is separated, the second step of the method is to separate local and internet traffic packets on the basis of packet protocol. Python script filters

out all Domain Name Server (DNS) packets from Internet traffic of each device separately. Afterwards, DNS packets are scanned and count both distinct domains and similar (repeated) domain packets and save both type of information for future use. IoT devices commonly transmit limited number of DNS packets and most of their packets are sent to the same domain, When we inspect IoT DNS packets in depth we found that most of the packets have same destination domain with same query and packet size, even some devices send similar packets in regular intervals on the same domain which clearly shows that IoT devices send limited DNS packets to only a few domains. We iteratively examined all devices separately and found from our dataset that IoT devices sent DNS queries maximum at 10 different domains. So, from this result, we set a threshold of DNS count equal to 10. On the other side, script count repeated packets of devices traffic by scanning destination Port, IP, Protocol Size of packets. If these features have common values then it shows that the device transmits repeated or similar packets. If the device contains at least 25% repeated packets, then there is more chance that the device is IoT. We get this value by iterative method by inspecting different devices traffic. We conclude that 25% is minimum required that IoT device repeat its packets. Then we recognize device traffic patterns, find common subsequences, how the device repeats its activity, after how much time or remain constant, also identify traffic flow a device sends and perceives the behavior of the device. We also show a visual presentation of the traffic flow of some devices by using the Origin tool listed in Table 2 which shows the understandable representation of device activity in a single graph. At the end of devices, profiling the K-Means clustering algorithm is applied on profile data to make clusters of these devices.

3. EXPERIMENTS AND RESULTS

During experiments our model identifies that the number of protocols used by IoT devices is much less as compared to non-IoT devices, as shown in Fig.7.

IoT devices are special-purpose devices, running limited services, and use fewer protocols. Non-IoT devices are generic and it depends on users how to configure their machine and allow services and protocol, commonly hundreds of applications are running on the machine and use a vast variety of protocols randomly. We identify number of protocols used by IoT devices that are in the range: 5-15, while Non-IoT devices use a minimum of 17 protocols in our dataset. Commonly, a desktop/laptop machine uses more than 30 protocols. DNS is one of the most common protocols used by almost every network device.

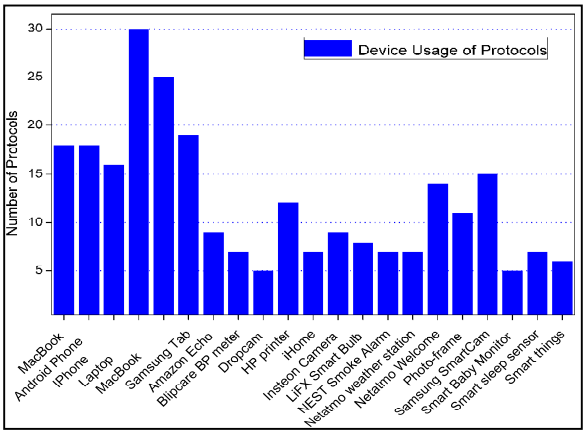

Fig. 7: Number of Protocols used by Devices

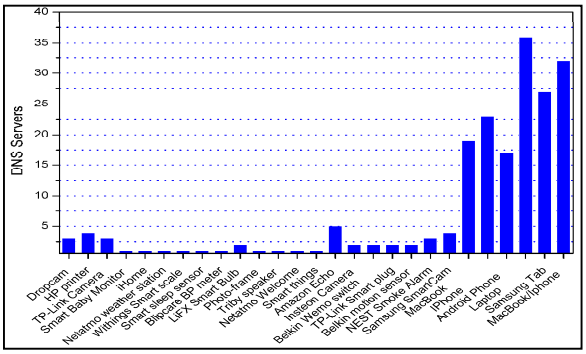

Fig. 8: Number of DNS servers used by a device

Some devices use DNS frequently whereas others utilize less DNS protocol, as shown in Fig. 8. IoT devices use DNS protocol very rarely with limited DNS servers, commonly 1-20 packets per day only which are clearly distinguishable by non-IoT devices, whereas non-IoT devices send hundreds of DNS queries randomly depending on users' interest and needs as listed in Table 3.

We examine mostly the IoT devices communicate with their vendor's domains for example devices made

by Belkin [26] sends DNS to Belkin.com shown in Table 3. It is observed that IoT based devices generated only 0.3% DNS packets from the whole dataset of traffic. Another example is Withings brand [26] devices. These devices send DNS query only to their own domain *withings.net* as shown in Table 3. All Withings band product accessed very less frequency while MacBook used DNS frequently, it visited multiple domain servers in a few hours, and URLs are listed in Table 3. which confirm that Non-IoT devices send more DNS as compared to IoT devices.

Table.3 DNS Server Summary of Devices

Device Name	Device Type	DNS Queries URL
Baby Monitor	IoT	babyws.withings.net
TP-Link Smart plug	IoT	devs.tplinkcloud.com uk.pool.ntp.org
iHome	IoT	api.evrythng.com
Weather station	IoT	netcom.netatmo.net
Smart scale	IoT	scalews.withings.net
Blipcare BP meter	IoT	tech.carematix.com
Triby speaker	IoT	sip.invoxia.com
Netatmo Welcome	IoT	apicom.netatmo.net
MacBook	Non - IoT	notify.dropbox.com www.apple.com, imap.gmail.com drive.google.com ax.itunes.apple.com www.adobe.com, talk.google.com client.dropbox.com outlook.office365.com accounts.google.com platform.twitter.com www.facebook.com play.google.com, docs.google.com mail.google.com

Like DNS, Network Time Protocol (NTP) is another most popular protocol used by IoT devices regularly to sync with their servers periodically. We examine some devices such as TP-link smart plug, Samsung smart camera, LIFX smart bulb, *etc*. They sync their time with publicly available server pool.ntp.org. We also identify some IoT devices which use NTP with a detectable pattern like smart thing and Belkin motion sensor sends NTP packet after every 10 minutes, LIFX sends NTP packet after every 5 minutes and TP-link Smart plug sends NTP packet every hour as shown in Fig. 9. EAPOL, TCP protocols were also used by most of the devices in regular patterns like smart thing send TCP packets of size 60 bytes after every 10 seconds to keep-alive information.

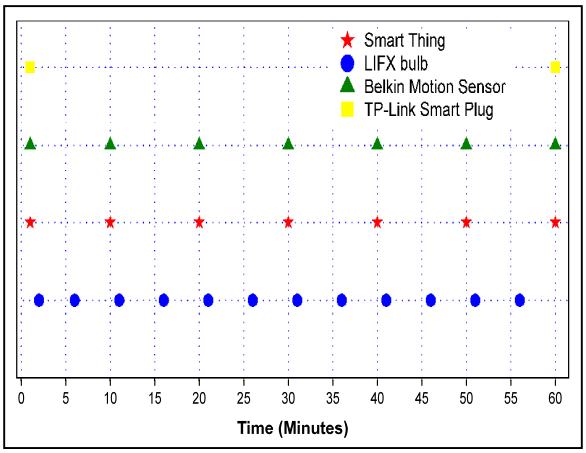

Fig. 9: NTP packets pattern of devices

IoT device behavior and traffic patterns have a huge impact on infrastructure planning, forecasting, network scaling, and support services. To understand the Network traffic pattern, we analyzed time-series data of some device traffic. Most of IoT devices make a frequent pattern that clearly shows the IoT devices generate periodic traffic.

Compared to other devices Nest protect smoke alarm [26]. The traffic rate was among those devices which generate 25 KB data traffic in 90 seconds with the repeated pattern as shown in Fig. 10. Nest protect smoke alarm sends 200 packets daily and the pattern of packets (traffic) is almost the same (up to 97 %). Nest smoke detector makes DNS request only 3 times in a day and communicates with 3 unique servers only as shown in Table 3.

Withings smart scale network traffic [26] generates traffic only when it is used by a user. Smart Scale generates cyclic network traffic which is clearly seen in Fig. 11. We found that it behaves in a periodic way of long transmission. We monitor that it generates half-cycle about 150 packets then generates repeated traffic

again like previous packets. Like other devices, the Blipcare BP meter [26] also generates similar traffic whenever in-use.

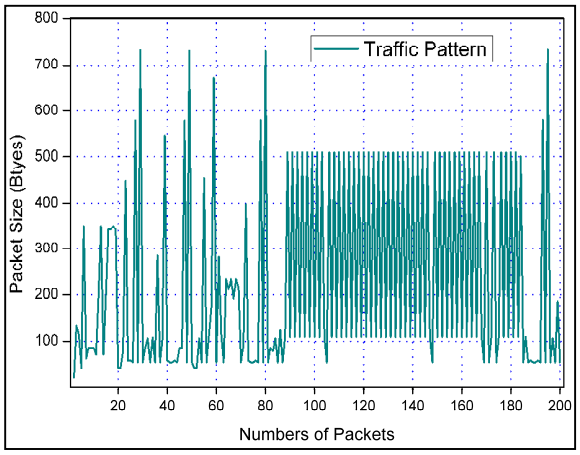

Fig. 10: NEST Smoke Alarm Traffic Pattern

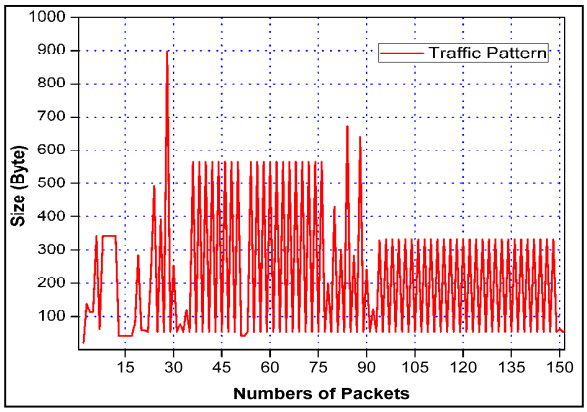

Fig. 11: Withings Smart Scale Activity Cycle Pattern

Device activity traffic is shown in Fig. 12. Device traffic contains 60 packets in a single activity, and the duration of traffic is less than one minute.

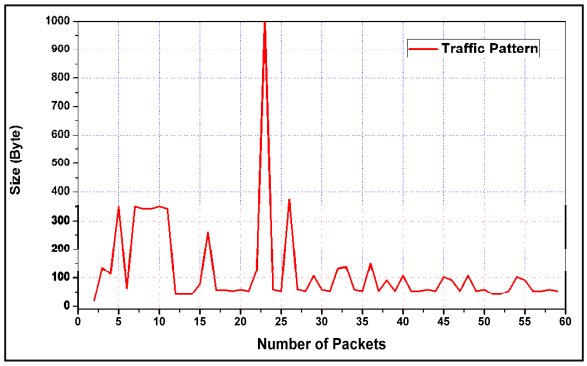

Fig. 12: BlipCare BP Meter Activity Cycle

3.1 Device Profiling

There are three steps to build a profile of devices.

(i) The user of a device makes a profile of the devices.
(ii) The manufacturer of a device makes a profile of their devices.
(iii) Automatic device profiling by scanning device traffic using the ML algorithms.

In this research, the first method is used for profiling in collaboration with Data mining techniques. We make an almost fine profile for most of the devices. By analyzing complete traces of the device and using device traffic pattern we make a device profile with respect to device behavior.

In order to explain the device profile which should be recognized, the profile must include information about how device responds, what protocols are running on the device, min, max, and average packet size, and number of ports used by devices. The number of devices uses DNS to resolve the IP endpoints on the Internet, DNS queries must also be considered in the profile. Port numbers and protocols used by IoT devices are fundamental elements that describe the device traffic insights, it will help to explore the device and make an effective and accurate device profile.

LIFX light bulb is another device traffic available in [26]. It is a type of light bulb which is controlled by a smartphone, it changes its color and light intensity, *etc*. After analyzing the wide-ranging behavior of device traffic, we make a profile. The summarized profile detail is described in Table 4.

Table 4**:** LIFX Smart Bulb Profile

Device Name	LiFX Smart Bulb
MAC Address	d0:73: d5:01:83:08
Connection Type	Wireless
Response Type	Event + Continuous
Traffic Duration	Continuous Active
DNS servers	1
Max Packet Size	784
Avg Packet Size	95
Min Packet Size	20
No of Ports	11
No of Protocols	9
Traffic Pattern	Constant

As discussed, earlier Nest protect smoke alarm is an IoT device. A brief profile is summarized in Table 5.

Table5: NEST Smoke Alarm Profile	
Device Name	**NEST Smoke Alarm**
MAC Address	18-b4-30-25-be-e4
Connection Type	Wireless
Response Type	Periodic
Traffic Duration	1 minute per day
DNS servers	2
Max Packet Size	732
Avg Packet Size	230
Min Packet Size	20
No of Ports	6
No of Protocols	8
Traffic Pattern	Repeated

Samsung smart camera another IoT device traffic is available on [26] considered in the current study. After examining device pattern and behavior we make a profile, summarized profile detail is described in Table 6**.**

Table 6: Samsung Smart Camera Profile	
Device Name	**Samsung Smart Camera**
MAC Address	00:16:6c: ab:6b:88
Connection Type	Wireless
Response Type	Continuous
Traffic Duration	Continuous
DNS servers	4
Max Packet Size	1514
Avg Packet Size	256
Min Packet Size	20
No of Ports	16
No of Protocols	14
Traffic Pattern	Random

3.2 IoT Device Clusters

The clustering of IoT devices is the final step of the experiment. After profiling devices, the clustering algorithm is used on profile data which in new approach in IoT categorization before that classification is used on device traffic data instead of device profile information. K-means Algorithm was applied in our dataset. We set the value of K equal to 3, which means algorithms makes three clusters. The first cluster grouped those devices which repeat their traffic patterns means those devices that repeatedly generate similar traffic in the activity cycle. Cluster 2 is composed of those devices which make semi pattern (generate repeated traffic on some protocols only and remaining traffic was random). Cluster 3 is for remaining devices that generate random traffic. K-mean uses the Euclidian distance (equation 1) for distance measurement. Euclidian is commonly used in many algorithms for similarity measurement, the detailed clusters are shown in Table 7.

$$\mathrm{dist}(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (1)$$

Table 7: Clusters of Device	
Cluster 1 Fully Repeated	Nest Smoke Alarm, BlipCare BP monitor, Withings Smart Scale, TP-link Smart plug.
Cluster 2 Semi Repeated	Withing baby monitor, Amazon Echo, Netatmo Weather station, Belkin motion sensor, Belkin switch, LIFX bulb
Cluster 3 Random Traffic	Insteon Camera, iHome, Triby Speaker, HP printer, PIX-Star photo Frame, Samsung smart Camera

4. CONCLUSIONS

Despite the proliferation of IoT devices in a smart environment, operator of such network lacks visibility into what IoT devices are connected to their network, what their traffic characteristics are, what is the pattern of activities and whether the devices are functioning properly without breach of organization policies. The purpose of our work is to show the behavior of IoT devices, identify device patterns on how devices send traffic, which protocols they used most and continuously. We found DNS queries patterns that were clearly different from non-IoT devices. The traffic pattern of some devices are constant which shows that when a user uses a device it generates the same activity cycle making it very predictable. Our results are better to understand the difference between IoT and Non-IoT network traffic which is beneficial for the network administrator to make a better-quality network policy in respect of security, routing and optimally allocation of the resources. The visual results clearly describe patterns of devices. There are still some deficiencies such as a multi-standard IoT devices, limited number of smart devices, few and common traffic traces of IoT devices are available, that is why researchers may not be able to make a

complete profile of devices by investigating all scenarios. Thus, there is a need to analyze a large number of IoT devices in different scenarios. There is a necessity for standardization, more accurate comprehensive profiling of IoT devices, architecture, and protocols to improve IoT network and overcome emerging challenges.

ACKNOWLEDGEMENT

This research is mainly based upon wok supported by the UIIT University of PMAS-Arid Agriculture Rawalpindi, Pakistan, under the supervision of Dr. Syed Mushhad Gilani, Assistant Professor (Computer Science), UIIT, PMAS-Arid Agriculture University Rawalpindi.

REFERENCES

1. Meidan Y,, Bohadana M., Shabtai A., Guarnizo J.D., Ochoa M., Tippenhauer N.O., Elovici Y., "ProfilIoT: A Machine Learning Approach for IoT Device Identification Based on Network Traffic Analysis", *Proceedings of the Symposium on Applied Computing (SAC'17).*, pp. 506–509, Marrakech, Morocco, 2017.
2. Aman A.H., Hassan R.A., Hashim A, Ramli H. A., "Investigation of Internet of Things Handover Process for Information Centric Networking and Proxy Mobile Internet Protocol", *Mehran University Research Journal of Engineering and Technology*, Vol. 38,No. 4, pp. 867–874, 2019.
3. Coetzee L., Eksteen J. "The Internet of Things – Promise for the Future ? An Introduction", *2011 IST-Africa Conference Proceedings*, pp. 1–9, Gaborone, Bostwana, 2011.
4. Parveen K., Ali A., Asadullah G., "Survey on Operating Systems for the Applications of the Internet of Things Introduction", *Journal of Information Communication Technologies and Robotic Applications*, Vol. 7, pp. 9–16, 2018.
5. Tsai C., Lai C., Chiang M., Yang L. T., "Data Mining for Internet of Things : A Survey", *IEEE Communications Survays and Tutorials,* Vol. 16, No. 1, pp. 77–97, 2014.
6. "IDC Forecast Report," 2018. [Online]. Available: https://www.forbes.com/sites/tomcoughlin/2018/11/27/175-zettabytes-by-2025/#f86f0255459. [Last accessed on 22nd August, 2020].
7. "IHS Scaling Report," 2019. [Online]. Available: https://statista.com/statistics/471264/iot-number-of-connected-devices-worldwide. [Last accessed on 1st July 2019]
8. Daud S., Gilani S. M. M., Hamid I., Kabir A., Nawaz Q., "Dsdv And Aodv Protocols Performance In Internet Of Things Based Agriculture System For The Wheat Crop Of Pakistan", *Pakistan Journal of Agricultural Aciences*, Vol. 57, No. X, pp. 1–9, 2020.
9. Ravidas S., Lekidis A., Paci F., Zannone N., "Access control in Internet-of-Things: A survey", *Journal of Network and Compututer Applications*, Vol. 144, No. June, pp. 79–101, 2019.
10. Cui L., Yang S., Chen F., Ming Z., Lu N., Qin J., "A survey on application of machine learning for Internet of Things", *International Journal of Machine Learning and Cybernetics*, Vol. 9, No. 8, pp. 1399–1417, 2018.
11. Wang Y., Xiang Y., Zhang J., Zhou W., Wei G., Yang L.T., "Internet Traffic Classification Using Constrained Clustering", *IEEE Transactions on Parallel and Distributed Systems*, Vol. 25, No. 11, pp. 2932–2943, 2014.
12. Miettinen M., Marchal S., Asokan N., "IOT SENTINEL Demo : Automated Device-Type Identification for Security Enforcement in IoT", *Proceedings of the 37th IEEE International Conference on Distributed Computing Systems*, pp. 2511-2514, Atlanta, GA, USA, 2017.
13. Hamza A., Ranathunga D., Gharakheili H. H., Roughan M., Sivaraman V., "Clear as MUD: Generating, validating and applying IoT behavioral profiles", *Proceedings of the 2018 Workshop on IoT Security and Privacy, Part SIGCOMM 2018*, pp. 8–14, Budapest, Hungary, 2018.
14. "IoT Scaling Forecast Report," 2018. [Online]. Available: https://www.ietfjournal.org/managing-the-internet-of-things-its-all-about-scaling. [Last accessed on 15th July 2019].
15. A. Sivanathan *et al.*, "Characterizing and classifying IoT traffic in smart cities and campuses", *Proceedings of the IEEE Conference on Computer Communications, (IEEE*

INFOCOM 2017), pp. 559–564, Atlanta, USA, May 2017.

16. Akram B.A., "Change Detection Algorithms for Surveillance in Visual IoT: A Comparative Study", *Mehran University Research Journal of Engineering and Technoly*, Vol. 37, No. 1, pp. 77–94, 2018.
17. "List of Smart Cities worldwide," 2019. [Online]. Available: https://www.nominet.uk/list-smart-city-projects/.
18. Rajput A. and K. Vinoth Babu, "Clustering techniques of Wireless sensor networks for Internet of Things", *ARPN Journal of Engineering and Applied Sciences*, Vol. 13, No. 5, pp. 1715–1733, 2018.
19. Shahid M. R., Blanc G., Zhang Z., Debar H., "IoT Devices Recognition Through Network Traffic Analysis", *Proceedings of the IEEE International Conference on Big Data,* pp. 5187–5192, Seattle, WA, USA, 2018.
20. Singhal P., Jodhpur S., "State of the Art Review of Network Traffic Classification based on Machine Learning Approach", *International Journal of Computer Applications,* Vol. 2013, pp. 12–15, 2013.
21. Sivanathan A., Sherratt D.,. Gharakheili H. H, Radford A., Wijenayake C., A. V., Sivaraman V., "Characterizing and Classifying IoT Traffic in Smart Cities and Campuses", *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 559-564, Atlanta, GA, USA, 2017.\
22. Wang Y., Xiang Y., Zhang J., Zhou W., Xie B., "Internet traffic clustering with side information", *Journal of Computer and System Sciences*, Vol. 80, No. 5, pp. 1021–1036, 2014.
23. Bai L., Yao L., Kanhere S. S., Wang X., Yang Z., "Automatic Device Classification from Network Traffic Streams of Internet of Things", *Proceedings of the 43rd IEEE Conference on Local Computer Networks*, pp. 1-9, Chicago, USA, 2018.
24. Amar Y., Mortier R., Brown A., Colley J., Crabtree A., "An Analysis of Home IoT Network Traffic and Behaviour", *arXiv:1803.05368v1 [cs.NI]* 14 Mar 2018.
25. Hamza A., Ranathunga D.,. Gharakheili H. H, Benson T. A., Roughan M., Sivaraman V., "Verifying and Monitoring IoTs Network Behavior using MUD Profiles", *IEEE Transactions on Dependable and Secure Computing*, 2020.
26. "Dataset link," 2017. [Online]. Available: https://iotanalytics.unsw.edu.au/iottraces.html.