ANALYSIS OF NETWORK TRAFFIC FEATURES GENERATED BY IoT DEVICES

Ivan Cvitić[1], Petra Zorić[1], Tibor Mijo Kuljanić[2], Mario Musa[3]
[1]University of Zagreb - Faculty of Transport and Traffic Sciences,
ivan.cvitic@fpz.hr, petra.zoric@fpz.hr
[2]HEP ODS d.o.o., Ulica grada Vukovara 37, 10 000 Zagreb
tibor.kuljanic@gmail.com
[3]Hrvatska Lutrija d.o.o., Ulica grada Vukovara 72, 10 000 Zagreb
mario.musa@lutrija.hr

Abstract: *Devices that coexist in the Internet of Things environment are growing in number and their application is becoming more diverse. This opens many problem areas for research. Examples of such areas are the classification of IoT devices, the detection of network traffic anomalies that such devices generate, and the monitoring and management of IoT devices and communication infrastructure. Certain researches indicate homogeneity of device behavior within individual groups of IoT devices and heterogeneity between different groups. The problem of defining groups of devices with similar characteristics is emphasized. This paper presents an analysis of network traffic features generated by IoT devices in order to gain insight into the traffic features that can be used as a framework for further research in a field of device class definition and device classification in the IoT environment.*

Key words: *classification, network traffic flow, anomaly detection, DDoS*

1. Introduction and previous research

The appearance of the Internet of Things (IoT) concept as a new direction in technological development and a new communication paradigm that brings together billions of new devices connected to the Internet, creates a new space for security vulnerabilities that can be exploited for unauthorized and malicious activities.

According to the forecasts presented in [1], until 2020, approximately 31 billion IoT devices will exist globally in use. In this case, 41% or 12.86 billion IoT devices will be installed within the concept of a smart home (SH) [2]. The limitations of IoT devices in general, and thus SHIoT (smart home IoT) devices, are described in a research [3], covering hardware constraints, high autonomy requirements and low production cost, which reduces the ability to implement advanced security methods and increases the risk of many threats presented in [4].

The traffic generated by SHIoT devices or MTC (Machine Type Communication) traffic is different from the traffic generated through conventional devices, HTC (Human

Type Communication) traffic, as shown by research [5]. Although SHIoT devices are characterized by heterogeneity, MTC traffic is homogeneous in contrast to HTC traffic, which means that devices of the same or similar purpose behave approximately equally, that is, generate traffic of similar characteristics [6], [7].

The specific features of MTC traffic have been used to solve many problems in the communication network. Research [8] looks at the impact of MTC traffic on QoS when integrating with HTC traffic in the LTE communication network. The identification and classification of IoT devices in smart cities and campuses, and smart environments using the characteristics of MTC traffic has been demonstrated by research [9] and [10]. Research [11] seeks to identify new requirements and challenges in the design and management of a mobile communications network imposed by the generation of MTC traffic.

This research aims to provide an overview of the traffic features that make it possible to differentiate a variety of smart home IoT devices as one of the fastest growing areas of the IoT concept application. As a result, a framework of relevant traffic features can be formed, according to which it is possible to distinguish IoT devices for further research in the field of classification of such devices and detection of their illegitimate behaviour.

2. The importance and representation of a smart home as an area of IoT concept application

According to Gartner, the largest representation and application of the IoT concept by the number of IoT devices used by 2017 was in the field of smart building environments. After 2017, the concept of a smart home is the environment that integrates the largest number of IoT devices [12]. The representation of IoT devices by application categories is shown in Figure 1, which shows the dominance of IoT devices in the private sector and implies a smart home environment, relative to business sectors.

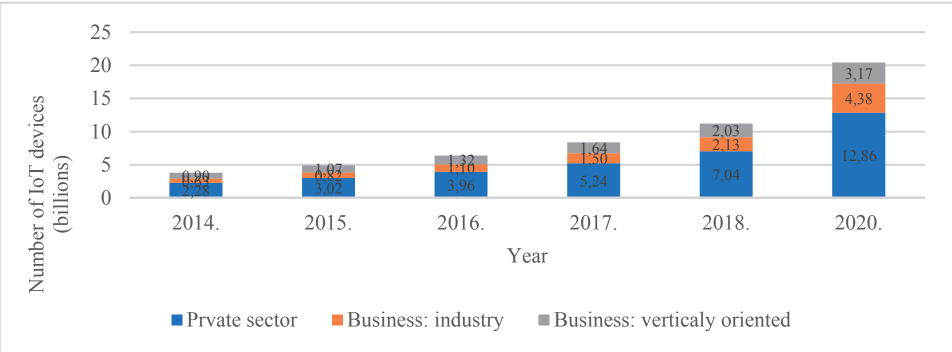

Figure 1. Number of implemented devices by category of application [1], [13]

A clearer insight into the representation of IoT devices by field of application is provided by IHS Markit survey [14]. Figure 2 shows that the smart home concept has the highest number of installed IoT devices (822.6 million) over other applications. The annual growth rate (prediction by 2021) is 19.6%, making the smart home concept, with the Industrial IoT concept (CAGR 23.4%) the fastest growing area of IoT concept application.

The number of smart homes that have implemented SHIoT devices from each category is shown in Figure 3. The figure shows a prediction of continued growth in device deployment across all these categories through 2023. According to [15], the largest increase is expected for homes with implemented SHIoT devices from the "monitoring and connectivity" category, which includes devices such as smart sockets, switches, and speakers. The statistical indicators presented in [16] indicate a continuous increase in revenue for this group of devices up to 2023 by region. The Asian (China) prediction indicates an annual revenue growth rate of 35%, while in the US and Europe it ranges from 17% - 25%.

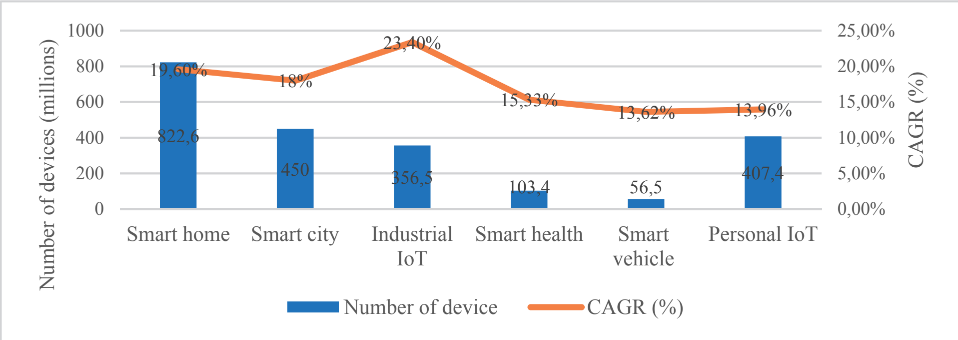

Figure 2. Number of IoT devices and annual growth rate by application area [14]

The second fastest growing smart homes are those that implement SHIoT devices from the "comfort and lighting" group, which includes devices such as lightning devices as the most common devices in this category, but also window and door sensors as well as controls such as garage doors management. Given the ease of implementation of devices in this category, which primarily applies to lightning devices, they often represent an entry point for users to implement smart home concepts. According to [17], the global market value of this group in 2023 will be approximately $14.32 billion. The expected annual growth rate of revenue for China is 41%, for Europe and the USA in the range of 19% - 27%.

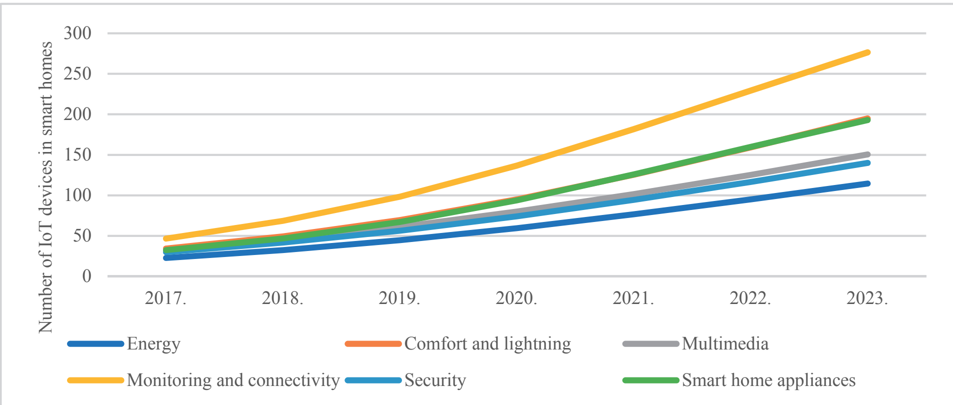

Figure 3. Number of smart homes with SHIoT devices implemented per category (2018-2023)[15]

According to the statistics presented, it is concluded that the number of devices in the IoT concept is growing exponentially. The application of the IoT concept dominates the private sector, that is, the smart home environment, as the area of application where most IoT devices are implemented. The number of IoT devices in the concept of smart home, with the area of industrial IoT, has the highest annual growth rate. The concept of smart home has a positive trend in terms of penetration of devices in the global market, the number of households within which SHIoT devices have been implemented, and market values regardless of the groups of devices combined under this concept. The indicators analysed accurately and unequivocally indicate that the smart home concept is currently the most represented and fastest growing area of application of the IoT concept.

3. Network traffic features of IoT devices in smart home environment

IoT traffic can be perceived as network activity through features such as traffic flow volume (sum of total downloaded traffic and total uploaded traffic), duration of traffic flow (time between the first and last packet in traffic flow), inactivity time of the device (time period in which the device does not have an active traffic flow). Network behavioural modelling is a commonly used approach to address challenges in the communication network, such as the detection of illegitimate traffic-based events generated by devices in the network. In general, current approaches seek to identify traffic characteristics at the network packet level and traffic flow level [18].

Numerous researchers are trying to identify the characteristics of traffic generated as a product of IoT device communication. The traffic characteristics generated by individual IoT devices can be a key factor in researching the causal relationships of generated traffic to certain processes in the communications network. Often, such features are used to identify IoT devices in the network [9], [10], [19], identification of the used type of services [20], detection of unauthorized devices in the network [21] and detection of network traffic anomalies [8], [22].

3.1 Network traffic features on network flow level

The traffic features that researchers observe depend on the goal of the research. Traffic intensity was used in [9] and [10] to distinguish between MTC and HTC traffic and the identification of IoT devices. Interpretation of the research results indicates that the traffic intensity generated by IoT devices is significantly lower (average 66 Kbps, peak 1 Mbps) than for conventional devices (average 400 Kbps, peak 17 Mbps for research) [9]. Differences between MTC and HTC traffic are also evident from the length of the session (95% of all IoT sessions observed lasts less than 5 seconds). The duration of a session also affects the amount of traffic transferred per session (in 75% of the sessions observed, the amount of traffic is less than 1KB, and in 1% of the sessions the amount of traffic is greater than 10 KB).

In addition to differentiating devices that generate MTC and HTC traffic by previous traffic characteristics, there is also a difference between individual devices or groups of devices that generate MTC traffic. According to research [10], individual IoT devices differ in the amount of traffic transmitted per traffic flow. For example, for LiFX smart lighting, the amount of data transmitted in most traffic flows is between 130 and 140 bytes, while for the Belkin motion sensor in most traffic flows, the amount of data transmitted is between 2800 and 3800 bytes. The same research also identified additional features that make it possible to differentiate individual IoT devices, such as data rates. So, in 60% of traffic, LiFX smart lighting transmits data at an average speed of 18 bps, while Belkin's motion sensor transmits data at a rate of 59-60 bps in 40% of traffic flows. The small amount of data transmitted throughout a traffic flow is evident in the same study whereby an analysis of this characteristic was performed at the level of individual devices. The same research also analyses the duration of traffic flow, where it was found that LiFX smart lighting generated most traffic flows (50%) in 60 seconds, while the Belkin motion sensor generates 21% of traffic flows in the same duration. Identified features on the traffic flow level are shown in Table 1.

Table 1. Traffic features generated by IoT devices at the traffic flow level

Label	Traffic feature	Feature explanation	Research
int_traff	Traffic intensity	The amount of traffic transferred per time	[9]
s_dur	Duration of the session	Time period in which the device generates traffic	[9]
sleep_time	Device idle time	The time period during which no active streams exist for the observed device	[9], [10]
flow_dur	Flow duration	The time period between the first and last traffic flow packets	[10]
flow_vol	Traffic volume	The total amount of incoming and outgoing traffic per traffic flow	[10]
avg_flow_rate	Average data rate of traffic flow	Ratio of traffic flow volume and flow duration	[10]
pack_size	Packet size	Packet size in traffic flow can be viewed through statistical measures such as mean, standard deviation, and minimum or maximum values	[8], [9], [19], [23]
proto	Protocols used	Communication protocols used in traffic flow	[10]
no_pack	Number of packets	The number of packets transmitted during the traffic flow	[8]
iat	Packet interarrival time	Time between the arrival of two consecutive packets in a traffic flow	[8], [23]

Research [19] seeks to classify IoT devices by semantic characteristics (IoT nodes, electronic devices, cameras, and switches) using traffic flow features such as packet length statistics, packet counts, and communication protocols used. The research assumes that all devices in a particular category have the same or approximately the same characteristics, which may not necessarily be true. Evidence of this is the 74.8% detection accuracy of the developed model. Packet size and packet interarrival time in traffic flow are also discussed by authors in research [23] who seek to identify IoT devices in a smart home environment.

3.2 Network traffic features on network packet level

Some studies, to identify IoT devices, detecting anomalies, or solving other class-oriented problems focused on considering the traffic features of such devices at the network packet level. Identified features on the network packet level are shown in Table 2. The research presented in [21] seeks to detect unauthorized IoT devices in a communications network. In doing so, three package-level features have been identified as relevant in the device classification. All three features relate to the TTL (Time to Live) of each package (minimum value, average value, and first quartile value).

Table 2. Traffic features generated by IoT devices at the packet level

Label	Traffic feature	Feature explanation	Research
proto	Presence of protocol	Monitoring the use of certain protocols in the current packet	[24], [25]
ttl	The number of network nodes the packet goes through	The value in an IP packet that tells network nodes whether to forward the packet to the next node or discard it	[21]
p_size	Packet size	Size of individually observed packet	[24]
ip_addr	Packet IP address	The source and destination IP address recorded in the packet header	[24]

Researches [23] and [25] use the network packet features generated by such devices on different TCP / IP layers to identify individual IoT devices. Feature values are binary, that is, indicate the presence of the observed feature such as IP addresses, source, and destination communication ports, use of certain protocols (ARP, LLC, IP, ICMP, HTTP, SSDP). Also features such as packet size, communication port class, and destination IP address counter are observed. Network packet features were also used in research [24] where features such as packet size, protocols, source, and destination IPs were observed for DDoS traffic detection.

4. Conclusion

Analysed research shows that traffic features are more frequently considered and used at the traffic flow level than at the network packet level. The aforementioned researches also use the features presented to identify individual devices or to classify them based on the semantic characteristics of the devices in question. Traffic flow as the level of observation and analysis of traffic features is selected because it represents the aggregated (statistical) data of the packet header for communication between source and destination. Packet-level traffic feature analysis captures more information such as package content, and also requires more computing resources to store and process. Given that most devices and applications nowadays use cryptographic methods when communicating, the contents of a packet cannot be viewed and analysed in an economically, timely and legally acceptable manner. Accordingly, observing and analysing traffic features at the traffic flow level is an acceptable and frequently used approach in numerous studies.

Future research of the problem area will seek to utilize the traffic features identified by this research for defining classes and developing a classification model of IoT devices in a smart home environment. The future research will aim to define classes of IoT devices and develop a classification model that will not depend on the semantic characteristics of the device or the individual device. Such an approach has the potential to be applied to currently existing IoT devices but also on future devices.

Literature

[1] Statista, "Internet of Things (IoT) connected devices installed base worldwide from 2015 to 2025 (in billions)," 2018. [Online]. Available: https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/. [Accessed: 24-Jun-2018].

[2] Statista, "The Internet of Things (IoT)* units installed base by category from 2014 to 2020 (in billions)," 2018. [Online]. Available: https://www.statista.com/statistics/370350/internet-of-things-installed-base-by-category/. [Accessed: 24-Jun-2018].

[3] I. Cvitić, M. Vujić, and S. Husnjak, "Classification of Security Risks in the IoT Environment," in *26-th Daaam International Symposium on Intelligent Manufacturing and Automation*, 2016, pp. 0731–0740.

[4] B. Ali and A. Awad, "Cyber and Physical Security Vulnerability Assessment for IoT-Based Smart Homes," *Sensors*, vol. 18, no. 3, p. 817, 2018.

[5] B. K. J. Al-Shammari, N. Al-Aboody, and H. S. Al-Raweshidy, "IoT Traffic Management and Integration in the QoS Supported Network," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 352–370, 2018.

[6] M. Laner, P. Svoboda, N. Nikaein, and M. Rupp, "Traffic models for machine type communications," in *10th IEEE International Symposium on Wireless Communication Systems 2013, ISWCS 2013*, 2013, vol. 9, pp. 651–655.

[7] I. Cvitić, D. Peraković, M. Periša, and M. Botica, "Smart Home IoT Traffic Characteristics as a Basis for DDoS Traffic Detection," in *Proceedings of the 3rd EAI International Conference on Management of Manufacturing Systems*, 2018, pp. 1–10.

[8] Y. Meidan *et al.*, "N-BaIoT: Network-based Detection of IoT Botnet Attacks Using Deep Autoencoders," *IEEE Pervasive Comput.*, vol. 13, no. 9, pp. 1–8, 2018.

[9] 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*, 2017, pp. 559–564.

[10] A. Sivanathan *et al.*, "Classifying IoT Devices in Smart Environments Using Network Traffic Characteristics," *IEEE Trans. Mob. Comput.*, vol. 18, no. 8, pp. 1745–1759, Aug. 2019.

[11] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang, "Large-scale measurement and characterization of cellular machine-to-machine traffic," *IEEE/ACM Trans. Netw.*, vol. 21, no. 6, pp. 1960–1973, 2013.

[12] S. Kejriwal and S. Mahajan, "Smart buildings: How IoT technology aims to add value for real estate companies," 2016.

[13] R. Meulen van der, "Gartner Says 8.4 Billion Connected "Things" Will Be in Use in 2017, Up 31 Percent From 2016." [Online]. Available: https://www.gartner.com/en/newsroom/press-releases/2017-02-07-gartner-says-8-billion-connected-things-will-be-in-use-in-2017-up-31-percent-from-2016. [Accessed: 12-Feb-2019].

[14] IHS, "The Internet of Things : a movement, not a market Start revolutionizing the competitive landscape," *IHS Markit*, 2017.
[15] "Smart Home - worldwide | Statista Market Forecast." [Online]. Available: https://www.statista.com/outlook/279/100/smart-home/worldwide. [Accessed: 09-Mar-2019].
[16] C. Blumtritt, "Smart Home Report 2019 – Control and Connectivity," Hamburg, 2019.
[17] C. Blumtritt, "Smart Home Report 2019 – Comfort and Lighting," Hamburg, 2019.
[18] D. Bekerman, B. Shapira, L. Rokach, and A. Bar, "Unknown malware detection using network traffic classification," in *2015 IEEE Conference on Communications and Network Security (CNS)*, 2015, pp. 134–142.
[19] L. Bai, L. Yao, S. S. Kanhere, X. Wang, and Z. Yang, "Automatic Device Classification from Network Traffic Streams of Internet of Things," *arXiv*, 2018.
[20] M. Lopez-Martin, B. Carro, A. Sanchez-Esguevillas, and J. Lloret, "Network Traffic Classifier With Convolutional and Recurrent Neural Networks for Internet of Things," *IEEE Access*, vol. 5, no. 1, pp. 18042–18050, 2017.
[21] Y. Meidan *et al.*, "Detection of Unauthorized IoT Devices Using Machine Learning Techniques," *arXiv*, 2017.
[22] I. Cvitić, D. Peraković, M. Periša, and M. Botica, "Novel approach for detection of IoT generated DDoS traffic," *Wirel. Networks*, vol. 25, pp. 1–4, 2019.
[23] M. Miettinen, S. Marchal, I. Hafeez, N. Asokan, A.-R. Sadeghi, and S. Tarkoma, "IoT SENTINEL: Automated Device-Type Identification for Security Enforcement in IoT," in *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, 2017, pp. 2177–2184.
[24] R. Doshi, N. Apthorpe, and N. Feamster, "Machine Learning DDoS Detection for Consumer Internet of Things Devices," in *2018 IEEE Security and Privacy Workshops (SPW)*, 2018, pp. 29–35.
[25] B. Bezawada, M. Bachani, J. Peterson, H. Shirazi, I. Ray, and I. Ray, "Behavioral Fingerprinting of IoT Devices," in *Proceedings of the 2018 Workshop on Attacks and Solutions in Hardware Security - ASHES '18*, 2018, pp. 41–50.

Rezime: *Broj uređaja koji koegzistiraju u Internet of Things (IoT) okruženju je u stalnom porastu, a njihova primena postaje sve raznovrsnija. To otvara mnoge istraživačke oblasti i probleme. Primeri takvih oblasti su klasifikacija IoT uređaja, otkrivanje poremećaja mrežnog saobraćaja kojeg generišu takvi uređaji, nadzor i upravljanje IoT uređajima i komunikacionom infrastrukturom. Pojedina istraživanja ukazuju na homogenost ponašanja uređaja unutar pojedinih grupa IoT uređaja i heterogenost između različitih grupa. Posebno se ističe problem definisanja grupa uređaja sa sličnim karakteristikama. U ovom radu analiziran je mrežni saobraćaj generisan IoT uređajima kako bi se dobio uvid u karakteristike saobraćaja koje se mogu koristiti kao okvir za dalja istraživanja u oblasti definicije klasa uređaja i klasifikacije uređaja u IoT okruženju.*

Ključne reči: *klasifikacija, tok mrežnog saobraćaja, otkrivanje poremećaja, DDoS*

ANALIZA KARAKTERISTIKA MREŽNOG SAOBRAĆAJA GENERISANOG IoT UREĐAJIMA

Ivan Cvitić, Petra Zorić, Tibor Mijo Kuljanić, Mario Musa