Generating IoT Edge Network Datasets based on the TON_IoT Telemetry Dataset

Georgios Zachos
Instituto de Telecomunicações
Aveiro, Portugal
Faculty of Engineering and Science
University of Greenwich
Chatham Maritime, UK
g.zachos@av.it.pt

Ismael Essop
Faculty of Engineering and Science
University of Greenwich
Chatham Maritime, UK
i.a.essop@greenwich.ac.uk

Georgios Mantas
Instituto de Telecomunicações
Aveiro, Portugal
Faculty of Engineering and Science
University of Greenwich
Chatham Maritime, UK
gimantas@av.it.pt

Kyriakos Porfyrakis
Faculty of Engineering and Science
University of Greenwich
Chatham Maritime, UK
k.porfyrakis@greenwich.ac.uk

José C. Ribeiro
Instituto de Telecomunicações
Aveiro, Portugal
jcarlosvgr@av.it.pt

Jonathan Rodriguez
Instituto de Telecomunicações
Aveiro, Portugal
Faculty of Computing, Engineering and Science,*University of South Wales*
Pontypridd, UK
jonathan@av.it.pt

***Abstract*—The rise of the Internet of Things (IoT) and Industrial IoT (IIoT), over the past few years, has been beneficial for the citizens, societies and industry. However, their resource-constrained and heterogenous nature renders them vulnerable to a wide range of threats. Therefore, novel security mechanisms, such as accurate and efficient anomaly-based intrusion detection systems (AIDSs), are required to be developed before IoT/IIoT networks reach their full potential in the market. However, there is a lack of up-to-date, representative and well-structured IoT/IIoT-specific datasets that are publicly available to the research community and constitute benchmark datasets for effective training and evaluation of Machine Learning models suitable for AIDSs in IoT/IIoT networks. Contribution to filling this research gap is of utmost importance and toward this direction the novel "TON_IoT Telemetry" dataset was recently published. Taking the opportunity to explore further this dataset, we targeted at its network-related part so as to generate IoT edge network specific datasets for effective development of more accurate and efficient IoT/IIoT-specific AIDSs. Therefore, in this paper, we present the methodology we followed to generate a set of IoT edge network specific datasets based on the "ToN_IoT Telemetry" dataset.**

Keywords—IoT cybersecurity, anomaly-based intrusion detection, dataset generation, record selection

I. INTRODUCTION

Over the past few years, Internet of things (IoT) and Industrial IoT (IIoT) networks have been bringing significant benefits to citizens, society and industry [1], [2]. However, the wide range of different communication technologies (e.g., WLANs, Bluetooth, Zigbee) and types of nodes/devices (e.g., sensors), incorporated in IoT/IIoT edge networks, are vulnerable to various types of security threats [3], [4], . In turn, this fact raises many security and privacy challenges for such networks as well as for the systems relying on these networks [2], [5], [6]. For instance, an attacker may compromise IoT/IIoT networks in order to manipulate sensing data (e.g., by injecting fake data) and cause malfunction to the IoT/IIoT-based systems that rely on the compromised IoT/IIoT networks. Thus, security solutions protecting IoT/IIoT networks from attackers are essential for the acceptance and wide adoption of such networks in the coming next years.

However, the high resource requirements of complex and heavyweight conventional security mechanisms cannot be afforded by (i) the resource-constrained IoT/IIoT nodes (e.g., sensors) with limited processing power, storage capacity, and battery life; and/or (ii) the constrained environment in which the IoT/IIoT nodes are deployed and interconnected using lightweight communication protocols [7]. Therefore, novel security mechanisms, such as accurate and efficient anomaly-based intrusion detection systems (AIDSs) adapted to the resource-constrained characteristics of IoT/IIoT networks, are necessary to be developed in order to address the pressing security challenges of IoT/IIoT networks with reasonable cost, in terms of processing and energy, before IoT/IIoT networks gain the trust of all involved stakeholders and reach their full potential in the market [1] , [2], [5].

AIDSs use Machine learning (ML) models and require appropriate benchmark datasets in order to be trained and evaluated [8], [9]. Nevertheless, there is a lack of up-to-date, representative and comprehensive IoT/IIoT-specific datasets that are publicly available to the research community and considered as benchmark datasets for effective training and evaluation of ML models suitable for AIDSs in IoT/IIoT networks. This lack of benchmark IoT/IIoT datasets constitutes a significant research challenge that should be addressed so as to develop more accurate and efficient IoT/IIoT-specific AIDS. Toward this direction, the authors in [10] have proposed, for the first time, to the best of our knowledge, a new dataset, called "TON_IoT Telemetry", which can be considered as a significant step toward a benchmark IoT/IIoT dataset, publicly available, for accurate design and evaluation of AIDSs as it includes events of a variety of IoT-related attacks and legitimate scenarios, IoT telemetry data collected from heterogeneous IoT/IIoT data sources, network traffic of IoT/IIoT network, and audit traces

The research work leading to this publication has received funding through the Moore4Medical project under grant agreement H2020-ECSEL-2019-IA-876190 within ECSEL JU in collaboration with the European Union's H2020 Framework Programme (H2020/2014-2020) and Fundação para a Ciência e Tecnologia (ECSEL/0006/2019).

of operating systems. Taking the opportunity to explore further the "TON_IoT Telemetry" dataset, we focused on the network-related part of this dataset in order to generate IoT edge network specific datasets for effective development of more accurate and efficient IoT/IIoT-specific AIDSs.

Therefore, in this paper, the main objective is the generation of a set of IoT edge network specific datasets based on the "Processed Network" datasets of the "ToN_IoT Telemetry" dataset [10]. In particular, from the network part of the "ToN_IoT Telemetry" dataset, we specifically filtered the records related to the "edge" network, as these records are the most appropriate for training and testing of ML models for AIDSs protecting IoT/IIoT networks, and generated a dataset, called "IoT Edge Network_Initial". However, the initially generated dataset (i.e., "IoT Edge Network_Initial") suffered from imbalances derived from the values of the "type" and "label" features. Therefore, we proposed a novel records selection algorithm in order to sub-sample the "IoT Edge Network_Initial" dataset and generate a new set of IoT edge network specific datasets where the imbalances of the initial generated dataset (i.e., "IoT Edge Network_Initial") were reduced.

Following the introduction, this paper is organized as follows. Section II reviews the testbed and the created "ToN_IoT Telemetry" datasets files. Section III presents the methodology followed to generate the set of IoT edge network specific datasets based on the "Processed Network" datasets of the "ToN_IoT Telemetry" dataset. Finally, Section IV concludes this paper and provides some hints for future work.

II. TON_IoT Telemetry dataset

In this subsection, the "TON_IoT Telemetry dataset" [10] is described. For the purposes of generation and collection of the TON_IoT Telemetry dataset [11], the authors in [10] developed a testbed integrating IoT sensors (e.g., weather and Modbus sensors), physical network components (e.g., switches, routers), several virtual machines (e.g., VMs of Offensive Kali systems, VMs of Windows client systems), hacking platforms, cloud and fog platforms and the devices were organized into the three layers of "Edge", "Fog" and "Cloud". Moreover, Software-defined Network (SDN) and Network Function Virtualisation (NFV) were employed through the NSX-VMware platform [12] in order to:

- establish both a virtualized "Fog" layer and a virtualized "Cloud" layer that simultaneously operate to offer the IoT/IIoT and network services;
- emulate and control multiple virtual machines (VMs) in the testbed for both hacking and normal operations; and
- manage the interaction between the three layers.

A. "Edge" layer of the Testbed

The "Edge" layer is fundamental in IoT/IIoT applications because its devices measure real-world physical conditions and transmit the collected information to the Fog or Cloud for further analysis [13]. The "Edge" layer of the testbed contains various IoT/IIoT devices, smartphones and smart TVs, physical gateways to the Internet as well as host systems. Additionally, the "Edge" layer incudes the physical host systems "NSX-VMware Server" and "vSphere System" which are used to deploy the fog layer and cloud layer, respectively, by means of virtualization through the NSX-VMware platform [12] and the NSX-VMware hypervisor platform respectively.

B. "Fog" layer of the Testbed

The purpose of the "Fog" layer is to extend the Cloud computing and services to the "Edge" layer of the network to provide limited computing capacity and storage near to the data sources [13]. The "Fog" layer of the testbed consists of VMs and the appropriate virtualization technology that manages the VMs and their services using the NSX-VMware platform. The included VMs and their roles are as follows:

- VMs where the Offensive Kali systems [14] are installed and include the scripts to simulate various attack scenarios;
- VMs (i.e., Metaspoitable3, OWASP security Shepherd, and Damn Vulnerable Web App (DVWA)) which offer vulnerabilities that can be exploited by the Offensive Kali systems [14];
- VMs of client systems (i.e., Windows 7 and 10);
- an Ubuntu 18.04 Middleware server where the Node-Red [15] and Mosquitee MQTT broker tools were deployed to manage the IoT/IIoT services and to operate seven IoT/IIoT sensors: weather, smart garage door, smart fridge, smart TCP/IP Modbus, GPS tracker, motion-enabled light, and smart thermostat;
- an Ubuntu 14.04 LTS orchestrated server that offered network services, including DNS (i.e., mydns.com), HTTP(s), DHCP, email server (i.e., Zimbra), Kerberos, and FTP, and generated network traffic between VMs; and
- a VM with the Security Onion tool that is used to log the network data of all the active systems in the testbed.

C. "Cloud" layer of the Testbed

The general purpose of the "Cloud" layer is to host large-size data centers with significant capacity for both computation power and storage to support IoT/IIoT applications and meet the resource requirements for big data analysis. The "Cloud" layer of the testbed includes:

- a Hive-MQTT broker [16] that is used to publish and subscribe the sensing data of the IoT/IIoT services using the Node-Red tool;
- a vulnerable PHP website [17] used to execute injection attacking events; and
- Cloud centers services (e.g., Microsoft Azure IoT Hub [18] and Amazon Web Services Lambda [19]) that were configured to subscribe and publish IoT/IIoT topics between them and the VMs of the "Fog" layer through the MQTT protocol.

D. ToN_IoT Dataset files

The authors in [10] simulated several different types of attack scenarios (i.e., Scanning, DoS, DDoS, ransomware, backdoor, data injection, Cross-site Scripting (XSS), password cracking and Man-in-The-Middle (MITM)) on their testbed and collected data from the different components of their testbed in dataset files. All the datasets are provided in files that follow the "csv" (comma separated vector) format. The datasets files

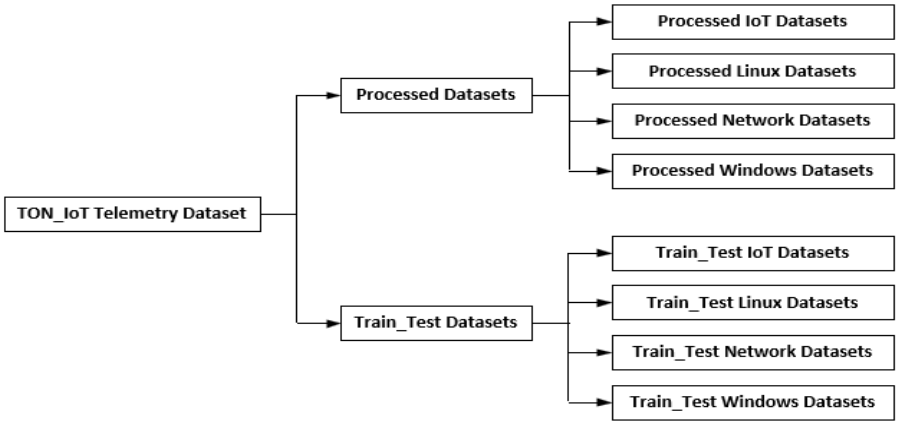

Fig. 1. ToN_IoT Telemetry datasets structure.

are split into two main folders: i) the "Processed" datasets folder, and ii) the "Train_Test" datasets folder. The "Processed" datasets contain a processed and filtered version of the datasets with: a) their standard features, b) a label feature (i.e., indicating whether an observation is normal or malicious), and c) a type feature (i.e., indicating the attacks sub-classes for multi-class classification problems) [10]. On the other hand, the "Train_Test" datasets contain selected records of the "Processed" datasets that were used by the authors in [10] as training and testing datasets for training and evaluating the accuracy and efficiency of various ML algorithms.

Both the "Processed" datasets and the "Train_Test" datasets consist of four types of dataset files (i.e., "Network", "IoT". "Linux", "Windows") with each referring to either the network traffic or a specific type of device (e.g., sensor, server, desktop) of the testbed, as it is also demonstrated in Figure 1. In particular, the "Network" datasets contain the traffic data that passed through the entire testbed and were captured during the simulations. The "IoT" datasets contain the data relating to each of the seven IoT/IIoT sensors that were simulated in the testbed. Finally, the "Linux" datasets and the "Windows" datasets contain the data relating to the two Ubuntu systems and the two Windows systems in the testbed, respectively.

III. "IoT EDGE NETWORK" DATASETS

A. *"IoT Edge Network_Initial" Dataset Generation*

In this work, the main objective is the generation of a set of IoT edge network specific datasets based on the "Processed Network" datasets of the "ToN_IoT Telemetry" dataset [10]. To this end, initially, we filtered the "Processed Network" datasets (i.e., csv files) and kept only the network records associated with the IoT edge layer of the testbed where the IoT devices are deployed. The filtering process was performed based on the sender node and the destination node of each record. Namely, a record was selected only if the sender node or the destination node existed in the edge layer. The created IoT edge network dataset was called as the "IoT Edge Network_Initial" dataset and was processed further, as described in section III.C, in order to create a set of specific and balanced "IoT Edge Network" datasets. Figure 2 summarizes the steps followed to generate a set of "IoT Edge Network" datasets based on the "Processed Network" datasets of the "ToN_IoT Telemetry" dataset.

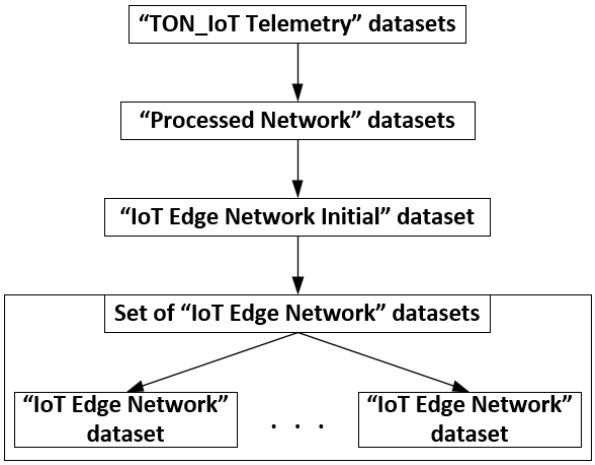

Fig. 2. Overiew of the steps followed to generate a set of "IoT Edge Network" datasets based on the "ToN_IoT Telemetry" dataset.

B. *"IoT Edge Network Initial" Dataset Analysis*

A statistical analysis is performed based on the "label" feature of the records on the "IoT Edge Network_Initial" dataset. The values of the "label" feature are "0" and "1" and indicate whether a record is categorized as normal or attack and thus, the values of the "label" feature correspond to a binary classification problem. Figure 3 shows the percentages of normal and attack records based on the total number of records in the "IoT Edge Network_Initial" dataset. An imbalance can be observed regarding the large amount of attack records compared to the normal records.

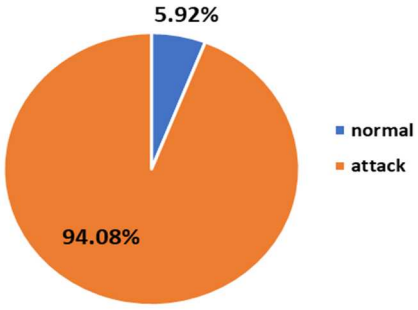

Fig. 3. Percentages of normal records and attacks records based on the total amount of records regarding the "IoT Edge Network_Initial" dataset.

Furthermore, we performed a statistical analysis based on the "type" feature of the records in the "IoT Edge Network_Initial" dataset. The unique values of the "type" feature in the "IoT Edge Network_Initial" dataset are 9 in total and they correspond to a classification problem with multiple classes. One of these values refers to the "normal" class and each of the remaining 8 values refers to a class related to a specific type of attack (Scanning, DoS, data injection, DDoS, password cracking, XSS, backdoor, MITM). Table I and Figure 4 shows the numbers and corresponding percentages of records belonging to the various classes of the dataset. It can be clearly seen that the records belonging to some attack classes (e.g., data injection, password, backdoor, MITM) are really few compared to the records related to other attack classes (e.g., Scanning, DDoS, XSS).

TABLE I. NUMBERS OF RECORDS REGARDING THE VARIOUS CLASSES OF THE "IOT EDGE NETWORK_INITIAL" DATASET.

Normal or Type of Attack	Number of records
Normal	203,277
Scanning	1,893,731
DoS	86,721
Data Injection	12,902
DDoS	630,141
Password Cracking	2,889
XSS	604,815
Backdoor	24
MITM	584
Total	**3,435,084**

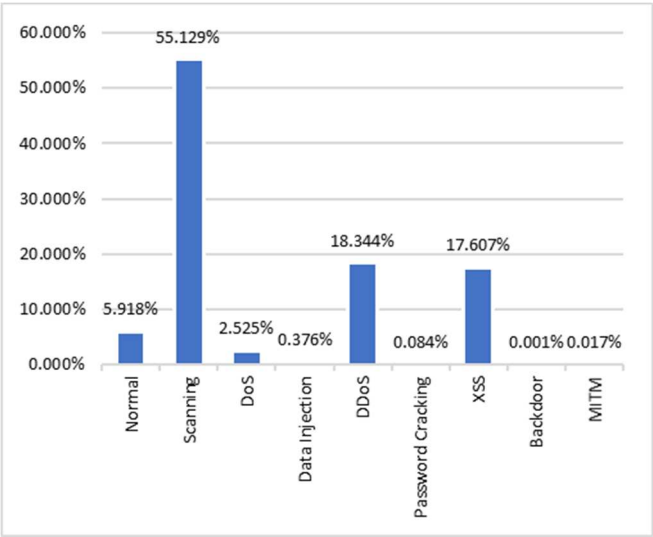

Fig. 4. Percentages of records regarding the various classes of the "IoT Edge Network_Initial" dataset.

C. *"IoT Edge Network" Datasets Generation*

Due to the imbalances related to the "label" feature and the "type" feature in the "IoT Edge Network_Initial" dataset, it is crucial to make use of an appropriate records selection algorithm (i.e., sub-sampling algorithm) in order to create a set of "IoT Edge Network" datasets, without the above mentioned imbalances, for effective training and evaluation of AIDSs suitable for IoT/IIoT networks.

We employed two different records selection algorithms and generated two types of training and testing datasets. Firstly, we followed a purely random sub-sampling algorithm and generated "random" training and testing datasets from the "IoT Edge Network_Initial" dataset. As it is also shown in Table II, in regards to the value of the "type" feature, the records of a created "random" dataset follow a value distribution similar to the "IoT Edge Network_Initial" dataset. This means that the percentage of the records belonging to a specific class in the "IoT Edge Network_Initial" dataset is almost equal to the percentage of the records belonging to the same class in the generated "random" dataset. At this point, it is essential to observe that only one record related to the "backdoor" class is included in the "random" dataset. This one record cannot be sufficient to train or test the performance of ML algorithms and thus, a purely random sub-sampling algorithm should not be considered as an appropriate algorithm in order to generate training and testing datasets from the "IoT Edge Network_Initial" dataset.

TABLE II. NUMBERS AND PERCENTAGES OF RECORDS REGARDING THE VARIOUS CLASSES OF THE FILTERED "IOT EDGE NETWORK_INITIAL" DATASET AND A GENERATED "RANDOM" SUBSET.

	"IoT Edge Network_Initial" dataset		**"random" subset**	
Value of "type" feature	***Number of records***	***Percentage of records***	***Number of records***	***Percentage of records***
Normal	203,277	5.918%	4,734	5.918%
Scanning	1,893,731	55.129%	44,103	55.129%
DoS	86,721	2.525%	2,020	2.525%
Data Injection	12,902	0.376%	300	0.375%
DDoS	630,141	18.344%	14,675	18.344%
Password Cracking	2,889	0.084%	67	0.084%
XSS	604,815	17.607%	14,086	17.608%
Backdoor	24	0.001%	1	0.001%
MITM	584	0.017%	14	0.018%
Total	**3,435,084**	**100.000%**	**80,000**	**100.000%**

Therefore, we propose a novel algorithm for records selection to reduce the above mentioned imbalances. The proposed algorithm has the following two objectives: (i) each attack class should have a similar amount of records to other attack classes, and (ii) the normal class should possess an amount of records that is similar to the total amount of attack records. The first objective focuses on balancing the records distributed among the attack classes, namely balancing the distribution of the "type" feature of the "IoT Edge Network_Initial" dataset, while the second objective focuses on balancing the records between the normal class and the attack classes and thus, the distribution of the "label" feature

of the "IoT Edge Network_Initial" dataset is balanced. The proposed algorithm is described below:

Algorithm: Records Selection
Input: Dataset, number of records to be selected (Nrecords), number of "attack" classes (NattackClasses)
Output: **"IoT Edge Network" dataset (out_dataset)**

1. *numNormalClasses* ← *1*
2. *classDiv* ← *numNormalClasses* * *NattackClasses* + *NattackClasses*
3. *idealNumRecordsPerAttClass* ← *round(Nrecords / classDiv)*
4. **For every** *class_i* **in** *(NattackClasses + numNormalClasses):*
5. *recordsIndexes_i* ← empty array
6. *selectedRecordsIndexes_i* ← empty array
7. **For every** *record* **in** *Dataset*:
8. *indexInDataset* ← *record.index()*
9. *class_i* ← *record.class()*
10. insert *indexInDataset* to the appropriate *recordsIndexes_i* array
11. *remClasses* ← *classDiv*
12. *remNumRecords* ← $N_{records}$
13. **For every** *class_i* **in** ($N_{attackClasses}$ + *numNormalClasses*):
14. **If** *class_i* is a "normal" class:
15. **If** *length(recordsIndexes_i)* ≤ $N_{attackClasses}$ * *idealNumRecordsPerAttClass*:
16. *remClasses* ← *remClasses* - $N_{attackClasses}$
17. *remNumRecords* ← *remNumRecords* – *length(recordsIndexes_i)*
18. **Else:**
19. **If** *length(recordsIndexes_i)* ≤ *idealNumRecordsPerAttClass*:
20. *remClasses* ← *remClasses* - 1
21. *remNumRecords* ← *remNumRecords* – *length(recordsIndexes_i)*
22. *numRecordsPerAttClass* ← *round*(*remNumRecords* / *remClasses*)
23. **For every** *class_i* **in** ($N_{attackClasses}$ + *numNormalClasses*):
24. **If** *class_i* is a "normal" class:
25. **If** *length(recordsIndexes_i)* ≤ $N_{attackClasses}$ * *numRecordsPerAttClass*:
26. *selectedRecordsIndexes_i* ← *recordsIndexes_i*
27. **Else:**
28. *selectedRecordsIndexes_i* ← random selection of $N_{attackClasses}$ * *numRecordsPerAttClass* records from the array *recordsIndexes_i*
29. **Else:**
30. **If** *length(recordsIndexes_i)* ≤ *numRecordsPerAttClass*:
31. *selectedRecordsIndexes_i* ← *recordsIndexes_i*
32. **Else:**
33. *selectedRecordsIndexes_i* ← random selection of *numRecordsPerAttClass* records from the array *recordsIndexes_i*
34. *datasetIndexes* ← empty array
35. **For every** *class_i* **in** *(NattackClasses + numNormalClasses):*
36. append *selectedRecordsIndexes_i* to array *datasetIndexes*
37. *Sort(datasetIndexes)*
38. *out_dataset* ← empty array
39. **For every** *index* **in** *datasetIndexes:*
40. *record_temp* ← Dataset[*index*]
41. insert *record_temp* to the *out_dataset* array
42. **Return** *out_dataset*

The proposed records selection algorithm is capable of creating a set of "IoT Edge Network" datasets from the "IoT Edge Network_Initial" dataset. Table III and Figure 5 show the numbers and corresponding percentages of records belonging to the various classes of a generated "IoT Edge Network" dataset. The records of the created dataset follow a more balanced value distribution, in terms of the "type" feature, compared to the "IoT Edge Network_Initial" dataset, shown in Figure 4.

TABLE III. NUMBERS OF RECORDS REGARDING THE VARIOUS CLASSES OF A GENERATED "IOT EDGE NETWORK" DATASET.

Normal or Type of Attack	Number of records
Normal	47,080
Scanning	5,885
DoS	5,885
Data Injection	5,885
DDoS	5,885
Password Cracking	2,889
XSS	5,885
Backdoor	24
MITM	584
Total	**80,002**

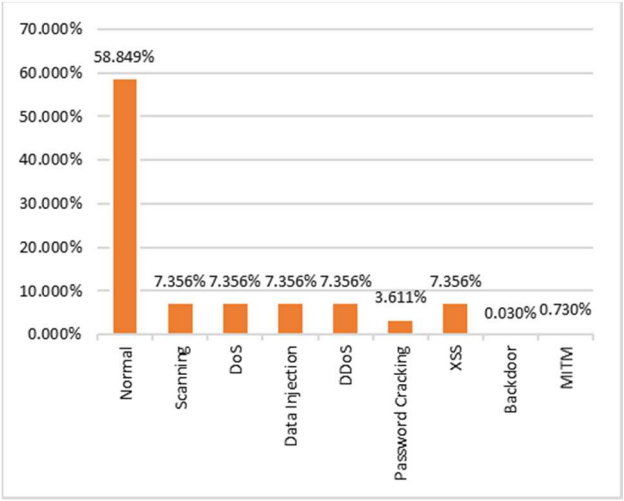

Fig. 5. Percentages of records regarding the various classes of a generated "IoT Edge Network" dataset.

IV. Conclusion

In this paper, we presented the methodology we followed to generate a set of IoT edge network specific datasets based on the "Processed Network" datasets of the "ToN_IoT Telemetry" dataset. In particular, as first step, we filtered the edge network part of the "ToN_IoT Telemetry" dataset and generated the "IoT Edge Network_Initial" dataset. However, we observed that the created dataset presented imbalances derived from the values of the "type" and "label" features. Therefore, we proposed a novel records selection algorithm in order to sub-sample the "IoT Edge Network_Initial" dataset and generate a new set of IoT edge network specific datasets where the imbalances of the initial generated dataset (i.e., "IoT Edge Network_Initial") were reduced. As future work, we plan to generate even more balanced and concise training and testing datasets from the edge network part of the "ToN_IoT Telemetry" dataset. Therefore, another algorithm that employs over-sampling techniques for the records of the classes with really few records will be designed and developed. Moreover, the importance of the features of the "IoT Edge Network_Initial" dataset should also be measured, using various ranking criteria, in order to generate more concise datsets using the proposed sub-sampling algorithm in this paper as well as the over-sampling algorithm planned as future work.

References

[1] L. da Xu, W. He, and S. Li, "Internet of things in industries: A survey," IEEE Transactions on Industrial Informatics, vol. 10, no. 4. IEEE Computer Society, pp. 2233–2243, Nov. 01, 2014. doi: 10.1109/TII.2014.2300753.

[2] E. Sisinni, A. Saifullah, S. Han, U. Jennehag, and M. Gidlund, "Industrial internet of things: Challenges, opportunities, and directions," IEEE Transactions on Industrial Informatics, vol. 14, no. 11, pp. 4724–4734, Nov. 2018, doi: 10.1109/TII.2018.2852491.

[3] M. Papaioannou et al., "A Survey on Security Threats and Countermeasures in Internet of Medical Things (IoMT)," Transactions on Emerging Telecommunications Technologies, p. e4049, 2020, doi: 10.1002/ett.4049.

[4] G. Mantas, N. Komninos, J. Rodriguez, E. Logota, and H. Marques, "Security for 5G Communications," in Fundamentals of *5G Mobile Networks*, wiley, 2015, pp. 207–220. doi: 10.1002/9781118867464.CH9.

[5] B. B. Zarpelão, R. S. Miani, C. T. Kawakani, and S. C. de Alvarenga, "A survey of intrusion detection in Internet of Things," *Journal of Network and Computer Applications*, vol. 84. Academic Press, pp. 25–37, Apr. 15, 2017. doi: 10.1016/j.jnca.2017.02.009.

[6] I. Makhdoom, M. Abolhasan, J. Lipman, R. P. Liu, and W. Ni, "Anatomy of Threats to the Internet of Things," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1636–1675, Apr. 2019, doi: 10.1109/COMST.2018.2874978.

[7] I. Essop, J. C. Ribeiro, M. Papaioannou, J. Rodriguez, G. Zachos, and G. Mantas, "Generating datasets for anomaly-based intrusion detection systems in iot and industrial iot networks," *Sensors*, vol. 21, no. 4, pp. 1–31, Feb. 2021, doi: 10.3390/s21041528.

[8] J. Ribeiro, F. B. Saghezchi, G. Mantas, J. Rodriguez, and R. A. Abd-Alhameed, "HIDROID: Prototyping a Behavioral Host-Based Intrusion Detection and Prevention System for Android," *IEEE Access*, vol. 8, pp. 23154–23168, 2020, doi: 10.1109/ACCESS.2020.2969626.

[9] J. Ribeiro, F. B. Saghezchi, G. Mantas, J. Rodriguez, S. J. Shepherd, and R. A. Abd-Alhameed, "An Autonomous Host-Based Intrusion Detection System for Android Mobile Devices," *Mobile Networks and Applications 2019 25:1*, vol. 25, no. 1, pp. 164–172, Feb. 2019, doi: 10.1007/S11036-019-01220-Y.

[10] A. Alsaedi, N. Moustafa, Z. Tari, A. Mahmood, and Adna N Anwar, "TON-IoT telemetry dataset: A new generation dataset of IoT and IIoT for data-driven intrusion detection systems," *IEEE Access*, vol. 8, pp. 165130–165150, 2020, doi: 10.1109/ACCESS.2020.3022862.

[11] Nour Moustafa, "ToN_IoT datasets," *IEEE Dataport*, Oct. 16, 2019. https://ieee-dataport.org/documents/toniot-datasets (accessed Jun. 08, 2021).

[12] "NSX-VMware." https://www.vmware.com/au/products/nsx.html (accessed Jun. 08, 2021).

[13] I. Stojmenovic and S. Wen, "The Fog computing paradigm: Scenarios and security issues," in *2014 Federated Conference on Computer Science and Information Systems, FedCSIS 2014*, Oct. 2014, pp. 1–8. doi: 10.15439/2014F503.

[14] "Kali Linux." https://www.kali.org/ (accessed Jun. 09, 2021).

[15] "Node-RED." https://nodered.org/ (accessed Jun. 09, 2021).

[16] "HiveMQ." https://www.hivemq.com/ (accessed Jun. 09, 2021).

[17] "Vulnerable Public PHP Website." http://testphp.vulnweb.com/ (accessed Jun. 09, 2021).

[18] "Microsoft Azure IoT Hub." https://azure.microsoft.com/en-au/services/iot-hub/ (accessed Jun. 09, 2021).

[19] "Amazon Web Services AWS Lambda." https://aws.amazon.com/lambda/ (accessed Jun. 09, 2021).