DOI: 10.15439/2023F3089
ISSN 2300-5963 ACSIS, Vol. 35

Cybersecurity Threat Detection in the Behavior of IoT Devices: Analysis of Data Mining Competition Results

Michał Czerwiński[†‡], Marcin Michalak[§‖], Piotr Biczyk[†**], Błażej Adamczyk[¶‖], Daniel Iwanicki[†], Iwona Kostorz[§],
Maksym Brzęczek[¶], Andrzej Janusz[†‡], Marek Hermansa[§], Łukasz Wawrowski[§], Artur Kozłowski[§]

[†]QED Software, Mazowiecka 11/49, 00-052 Warsaw, Poland
Email: {michal.czerwinski, piotr.biczyk, daniel.iwanicki, andrzej.janusz}@qed.pl
[§]Łukasiewicz Research Network - Institute of Innovative Technologies EMAG, Leopolda 31, 40-189 Katowice, Poland
Email: {marcin.michalak,iwona.kostorz,marek.hermansa,lukasz.wawrowski,artur.kozlowski}@emag.lukasiewicz.gov.pl
[¶]EFIGO sp. z o.o., M.Kopernika 8/6 , 40-064 Katowice, Poland
Email: {blazej.adamczyk, maksym.brzeczek}@efigo.pl
[‡]Faculty of Mathematics, University of Warsaw, Banacha 2, 02-097 Warsaw, Poland
Email: {m.czerwinski4}@uw.edu.pl, {a.janusz}@mimuw.edu.pl
{[‖]Department of Computer Networks and Systems, ** Faculty of Automatic Control, Electronics and Computer Science}
Silesian<file://Silesian> University of Technology, Akademicka 16, 44-100 Gliwice, Poland
Email: {marcin.michalak, blazej.adamczyk}@polsl.pl, pbiczyk@gmail.com<mailto:pbiczyk@gmail.com>

***Abstract*—The paper discusses a data science competition centered around the development of an anomaly detection system for IoT devices. The competition utilized a unique environment that allowed for the operation and monitoring of real IoT devices, including scheduling of attacks on these devices. The environment was used to collect the data, which included both normal and attack-induced behavior of IoT devices. The paper presents the background of the competition, the top models submitted, and the competition results. The paper also includes a discussion about restrictions related to the use of synthetic attack data as input for constructing anomaly detection systems.**

***Index Terms*—data science competitions; KnowledgePit.ai platform; cybersecurity; ML applications in log analysis; ML data quality**

I. Introduction

THE INCREASING number of Internet of Things (IoT) devices being used in present times implies the need to pay attention to ensure a proper level of their safety. The market review indicates that there is a lack of products that can increase the security of IoT devices while reducing the risk of successful attacks. This conclusion led to the idea of an IoT-dedicated system for detecting anomalies, which could be the result of an attack.

The consortium of EMAG, QED, and EFIGO runs a project focused on IoT devices cybersecurity – SPINET. Within that project, an environment that allows running and monitoring real IoT devices as well as collecting data describing their behavior has been developed. The environment also offers the possibility to schedule and perform attacks on monitored devices. The collected data, which describes both normal and attack-induced behavior of IoT devices, became the basis of the FedCSIS 2023 challenge.

This paper briefly presents the background of the competition, showcases the best models submitted to the challenge, and discusses the competition results and their potential impact on further system development.

II. Related Literature

Anomaly detection is a well–known approach for data analysis in many specific domains. As the IoT issues are becoming more and more interesting it is intuitive that any new or improved models should be tested on some data with anomalies to evaluate their capabilities. During the last decades, dozens of datasets related to network traffic security, operating systems or IoT monitoring were published. A brief summary is presented in Table I.

To reflect the nowadays trends in the data, we limited our search to datasets not older than 6 years and closely related to the IoT domain. Their short descriptions are presented below.

The environment for Bot–IoT [4] data capturing consisted of three components: network platforms, simulated IoT services, and extracting features. The network platforms included normal and attacking virtual machines. The IoT services simulating various IoT sensors were connected to the public IoT hub. The network environment that the dataset was collected from contained a combination of normal and botnet traffic. The dataset provides original packet capture (PCAP) files, generated Argus files and CSV files. The files separation is based on attack categories and subcategories. The dataset

TABLE I
DATASETS RELATED TO NETWORK TRAFFIC, OPERATING SYSTEM OR IOT SECURITY MONITORING (N- NETWORK, OS - OPERATING SYSTEM, IOT - INTERNET OF THINGS/INDUSTRIAL INTERNET OF THINGS DEVICES)

Dataset name	Owner	Monitoring	Reference
ADFA-LD	University of New South Wales	N, OS	[1]
Aposemat IoT–23	Stratosphere Laboratory	N, OS, IoT	[2]
CAIDA	Center of Applied Internet Data Analysis	N	[3]
Bot–IoT	University of New South Wales	N, IoT	[4]
CDX	United State Military Academy	N	[5]
DARPA 98-99	MIT Lincoln Laboratory	N, OS	[6]
KDD Cup 1999	University of California	N, OS	[7]
IoT Botnet	Ontario Tech University	N, OS, IoT	[8]
ISCX2012	University of New Brunswick	N	[9]
Kyoto	Kyoto University	N	[10]
Malware on IoT	Stratosphere Laboratory	N, OS, IoT	[2]
NSL-KDD	Canadian Institute for Cyersecurity	N, OS	[11]
RegSOC	Łukasiewicz–EMAG	N	[12], [13]
TON IoT	University of New South Wales	N, OS, IoT	[14]
Twente	University of Twente	N	[15]
UNSW-NB15	University of New South Wales	N	[16]
UMASS	University of Massachusetts	N	[17]
WUSTL-IIOT-2021	Washington University in St. Louis	N, IoT	[18]
Edge-IIoTset	Guelma Univ., De Montfort Univ., Annaba Univ., Edith Cowan Univ.	N, IoT	[19]

files include denial-of-service (DoS) and distributed denial-of-service (DDoS) attacks, operating system and service scanning, keyloggers and data exfiltration.

The TON_IoT datasets [14] are IoT and Industrial IoT (IIoT) datasets which files contain heterogeneous data sources collected from IoT and IIoT sensor telemetry data sets, Windows 7 and 10 operating systems datasets, as well as Ubuntu 14 and 18 Transport Layer Security (TLS) and network traffic datasets. The data was collected in a realistic and large-scale network. A testbed network was created for the Industry 4.0 network, which includes IoT and IIoT networks. The test platform was deployed using multiple virtual machines and hosts of Windows, Linux, and Kali operating systems to manage connections between the three tiers of IoT, Cloud, and Edge / Fog. Various attack techniques such as DoS, DDoS, and ransomware targeting web applications, IoT gateways, and computer systems on the IoT/IIoT network were conducted. The datasets were collected in parallel processing to collect several normal and cyber-attack events from network traffic, Windows audit trail, Linux audit trail, and IoT telemetry data.

The IoT-23 [20] is a dataset of network traffic from Internet of Things devices. The dataset consists of 23 captured different IoT network traffic scenarios. These scenarios were divided into twenty network captures from infected IoT devices that the malware samples were performed in each scenario and three network captures of the actual network traffic of the IoT devices. In each malicious scenario a specific malware sample was run on a Raspberry Pi. Scenarios included the following malware samples used to infect the device (Mirai, Torii, Trojan, Gagfyt, Kenjiro, Okiru, Hakai, IRCBot, Linux Mirai, Linux Hajime, Muhstik, Hide and Seek).

Malware on IoT [2] is a dataset of the monitoring of real IoT devices infected by malware. The dataset consists of labeled network traffic files stored during the long-lived real IoT malware traffic. It is divided into five subsets containing results of network traffic capturing during the Mirai malware attack and two subsets of honeypot network traffic capturing logs including protocols (HTTP, SSL, TCP, UDP) and connections statistics. The honeypot was a network camera.

The WUSTL-IIOT-2021 dataset [18] contains network data from industrial Internet of Things (IIoT) monitoring. The dataset was developed on test bench involving the industrial control systems (ICS) model including supervisory control and data acquisition (SCADA) systems. The testbed was dedicated to controlling a water storage tank, which is a part of the process of water treatment and distribution. The dataset was preprocessed and cleaned (rows with missing or corrupted values and extreme outliers removed. Artificially generated Command Injection, reconnaissance and DoS attacks were recorded in the set. It accounted for about 8% of network traffic.

The Edge-IIoTset is a dataset [19] containing monitoring data from IoT devices and IIoT applications. The IoT data was generated from more than 10 types of IoT devices such as low-cost digital temperature and humidity sensors, ultrasonic sensors, water sensors, level detection sensors, pH meters, soil moisture sensors, heart rate sensors, flame sensors, etc. Fourteen attacks related to IoT and IIoT communication protocols were identified and analyzed (divided into five threats) including DoS/DDoS attacks, information gathering, man-in-the-middle, injection attacks, and malware attacks.

The number of available datasets from IoT and IIoT devices monitoring is still relatively small in comparison to the rapidly growing number of such devices in the world (it is estimated that this year, their number will reach approx. 17 billion devices). Most of the available datasets contain data from network monitoring during normal operations and attacks. The data sets described in detail contain data from the audit of real and simulated IoT devices and their network environment. The available datasets providing kernel event monitoring data are from Solaris (DARPA 98-99).

It is also worth noting that this data science competition is the second cybersecurity-related challenge organized on the KnowledgePit.ai platform [21]. The first one, IEEE BigData 2019 Cup: Suspicious Network Event Recognition, was organized in 2019, jointly with Security-on-Demand company [22]. The platform also hosted a competition related to the monitoring of network devices [23]. Furthermore, KnowledgePit.ai has been a host to a number of data science competitions related to monitoring hazardous environments using networks of sensors [24].

III. FedCSIS 2023 Challenge

The challenge data were generated within the simulation environment which was an extension of the software framework described in detail in [25]. This real components simulation environment consisted of an IoT device (Raspberry Pi) and additional devices responsible for HTTP traffic generating and performing the attacks. The whole environment was plugged into a separated network to assure no other influences on the monitoring device.

It was necessary to model a normal way of device operation as well as to simulate some attacks to assure that the collected dataset contains both, safe and unsafe states of device usage.

In order to obtain a data sample describing both the normal operations of devices and the moments when attacks were carried out, it was necessary to ensure typical network traffic and triggering processes characteristic for it, as well as to prepare a scenario of external attacks on the device.

Typical operation conditions were generated continuously in several independent ways:

- SSH sessions: with the interval from 10 seconds up to 12 hours an administrator logs into the device and runs from 3 up to 10 commands (the time between each command varies from 0.5 to 11.5 seconds), later the administrator logs off;
- HTTP WAN traffic: the device has a built-in HTTP server, so it was possible to simulate cyclic queries; queries based on the real (other) WAN-connected device and their intervals were also taken from the historic data;
- file transfer: the file transfer service was run on the device (the endpoint) to simulate a periodic software update: a binary file of a size varying from 512 to 1,024 bytes was sent with the random interval from 1 to 12 hours;
- specialized HTTP queries: the device contained a dedicated endpoint for outer status checking/device clock synchronizing so it was possible to send the query that implied the "date -date now" command run (such a query was released with 9.5–11.5 second interval).

The environment provided the ability to perform two kinds of attacks: remote code execution and path traversal. In the case of the first one, the attack is carried out through a vulnerable endpoint "clock.php" and a query that uses a command injection vulnerability is invoked. Then, a reverse connection (with the attacking host) is established and an interactive session of the console "sh" program is run. Afterward, random commands are invoked with an interval of up to 20 minutes.

A path traversal vulnerability is used to upload the file into an unusual location (path) on the device. Files were saved into one of the following locations:

- /dev/shm/
- /var/tmp/
- /tmp/

The name of the file was random, as well as its size (from 20 to 5,024 bytes). Also, the number of files was varying (from 1 to 10) and the time between uploads was between 0.5 and 10 seconds.

A. Data preparation

System logs of each device were extracted, saved, and preprocessed resulting in a tabular dataset consisting of statistical characteristics of each feature aggregated over a rolling window of a fixed size.

The data created in such a way had certain characteristics typical of simulated data:

- The generated dataset contained a huge amount of information. Within this data, only a small fraction was collected during attacks on IoT devices. This resulted in big files which were hard to operate on containing only a small amount of data that could serve as valuable training data.
- Because the number of continuous attacks was small (not exceeding 20) it is reasonable to assume that the dataset makes it impossible to train a general-purpose IoT-attack-detection model. The methods chosen for generating attacks represent only a small fraction of the attack classes identified in the wild [26].
- Most of the created data was highly repeatable, resulting in a dataset of low diversity. This is normal behavior for IoT devices that operate in a repetitive manner.
- The training and testing data were created from a single source. This made it possible to achieve a near 100% accuracy on the testing data by identifying the process id (PID) values of processes that were targeted during an attack and using this knowledge to identify malicious activities in system logs. This is a highly improbable scenario in reality since restarting a process (or restarting the whole system) results in a new PID being assigned to the processes.

B. Evaluation procedure

The task in this challenge was to design an accurate method to predict whether system logs from an IoT system indicate the occurrence of cyberattacks or not. The quality of submissions was evaluated using the ROC AUC measure. The solutions were evaluated online and the preliminary results were published on the public leaderboard. The preliminary score was calculated on a small subset of test records, which was the same for all participants. The final evaluation was conducted after the completion of the competition using the remaining portion of the test records.

C. A baseline solution

Two baseline scores were established. The first one assumes a realistic scenario, while the second one was tailored to the dataset used in this competition, leveraging the knowledge that classification of PIDs enabled achieving a near 100% accuracy score on a portion of the test dataset.

1) Baseline score 1: The first baseline was calculated using an XGBoost model. Since the problem is a highly unbalanced classification task, the XGBoost's prior score was set according to the proportion of the system logs containing attacks (≈0.97).

The XGBoost model in this scenario was selected after comparing its results to Random Forest models (with and without class weights according to the proportion of the system logs containing attacks) and an XGBoost model without a prior score.

The baseline score achieved this way was 0.691 (ROC AUC). Examining the feature importances revealed that over 65% of the result was dependent on features created using the 'SYSCALL_pid' column which led to investigating the PID-based dependencies in the data and creating another classifier which gave the second baseline score.

2) Baseline score 2: Since the data used in this experiment was generated from a single artificial source, the PIDs corresponding to attacks were constant over time. For this reason, it is possible to list the PIDs of processes present during attacks and classify them in the test dataset as attacks.

This technique can also be altered by not strictly looking for all malicious PIDs in the test dataset but looking for PIDs that frequently occurred during attacks. Such an approach makes it possible to introduce a margin of error and thus filter out PIDs that could be falsely labeled in the training dataset as being part of an attack.

Performing a search-based classification as described above without any ML model resulted in a ≈0.93 ROC AUC score. Since a frequency-based method of classifying malicious PIDs was used; this score can be easily improved by further examining the PID values distribution in the training dataset.

IV. Challenge Outcomes

The competition was quite successful, with 78 participating teams and nearly 600 correctly formatted submissions. The majority of submitted solutions follow a general pattern of processing/cleaning the data → performing feature engineering → feature selection → model construction. However, there were some differences in the approach due to the hierarchical/complex form/format of the dataset. The internal data structure, i.e., a single observation is given as a separate file with a variable number of entries imposed the need for some form of aggregation. Among the submissions, we could observe different approaches in this regard. Some of the teams aggregated each of the input data files resulting in a representation where each observation (each file) was represented as a single row, while other teams concatenated all input files performing the aggregation only at the very end on the basis of predictions of classifiers working the level of file entries. The feature extraction/engineering stage was also approached differently by different teams. The total number of constructed features ranged from as few as several to several hundred thousand (including binary-encoded features). The most popular machine learning models used among the contestants fall into the category of gradient boosting machines - with particular implementations provided by commonly used open-source libraries like XGBoost, LightGBM, and CatBoost. However, also several other models could be encountered, including classical ones like decision trees, random forest, kNN, and logistic regression, as well as, custom methods, e.g., using micro-predictors build on top of features constructed using target guided binning, which achieved one of the best final scores.

TABLE II
Final results of the competition. The scores of the top 10 teams and their number of submitted solutions are shown.

Rank	Team name	Preliminary	Final score	#subs
1	MathLogic	1.0000	0.9999	76
2	dymitr	1.0000	0.9997	59
3	The Fellowship of the Cybersecurity	0.9997	0.9995	5
4	DML	0.9999	0.9993	176
5	Y-Team	1.0000	0.9986	8
6	Cyan	0.9940	0.9966	69
7	PisaTeam	1.0000	0.9957	10
8	hieuvq	0.9772	0.9718	101
9	Stan	0.9190	0.9293	14
10	baseline	0.9633	0.9257	-
...	...	...	...	...

Most of the competition participants decided to use the PID analysis-based approach to solve the task due to its high effectiveness. In such a case, the most significant differentiating factor between solutions from different teams was their approach to feature engineering. The final results of the top 10 teams are shown in Table II.

V. Conclusions and Future Work

The method of constructing the simulation environment and creating the competition dataset was sufficient to train and compare various machine learning models, but only to determine potential directions for future research and development specifically for cybersecurity data from IoT devices. It should be noted that such data is not suitable for training general-purpose machine learning models.

Based on the results of the competition and techniques employed, we plan future work to focus on the aspect of training data sets quality. Specifically - on tuning techniques for creating synthetic data sets that reflect various characteristics of real-life data. We plan to explore a hybrid approach, in which such synthetic data is used to augment data gathered from production IoT systems, in such a way as to create training sets that are optimal for training of a production system for analysis of anomalies in IoT behavior.

The end goal is to create a system that will be usable in various fields of application - most notably one that works on data gathered from utility providers (electricity, gas, water), manufacturers of video surveillance devices, and smart city infrastructure (interactive road signs, passenger information systems, control systems). Those companies will receive a toolkit that can be implemented in their own products. Additionally, the solutions can be utilized for protecting home devices such as smart lighting, electrical installations, alarm systems, and others.

VI. Acknowledgements

This work is a part of the "Development of a tool for protecting users, systems and devices of the Internet of Things, based on machine learning and behavioral analysis." project, which was co-financed by the Polish National Centre for Research and Development as part of the fourth CyberSecIdent–Cybersecurity and e–Identity competition (agreement number: CYBERSECIDENT/489240/IV/NCBR/2021).

References

[1] Creech G. and Hu J., "Generation of a new IDS test dataset: Time to retire the KDD collection," in *2013 IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2013, pp. 4487–4492, ISSN: 1558-2612.

[2] Stratosphere, "Stratosphere laboratory datasets," 2020, Retrieved March 15, 2021, from https://www.stratosphereips.org/datasets-overview.

[3] CAIDA, "Center of applied internet data analysis," 1998-2013, Retrieved March 16, 2021, from https://www.caida.org/catalog/datasets/completed-datasets/.

[4] Koroniotis N., Moustafa N., Sitnikova E., and Turnbull B., "Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-iot dataset," 2018.

[5] Sangster B., O'Connor T. J., Cook T., Fanelli R., Dean E., Adams W. J., Morrell C., and Conti G., "Toward instrumenting network warfare competitions to generate labeled datasets," in *Proceedings of the 2nd Conference on Cyber Security Experimentation and Test*, USA, 2009, CSET'09, p. 9, USENIX Association.

[6] MIT Lincoln Laboratory MIT, "Mit lincoln laboratory - darpa datasets," 1998-1999, Retrieved March 16, 2021, from https://www.ll.mit.edu/r-d/datasets.

[7] Stolfo S.J., Fan W., Lee W., Prodromidis A., and Chan P.K., "Cost-based modeling for fraud and intrusion detection: results from the jam project," in *Proceedings DARPA Information Survivability Conference and Exposition. DISCEX'00*, 2000, vol. 2, pp. 130–144 vol.2.

[8] Ullah I. and Mahmoud Q. H., "A technique for generating a botnet dataset for anomalous activity detection in iot networks," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2020, pp. 134–140.

[9] Shiravi A., Shiravi H., Tavallaee M., and Ghorbani A. A., "Toward developing a systematic approach to generate benchmark datasets for intrusion detection," *Computers & Security*, vol. 31, no. 3, pp. 357–374, 2012.

[10] Kyoto University, "Traffic data from kyoto university's honeypots," 2015, Retrieved March 17, 2021, from https://www.takakura.com/Kyoto_data/.

[11] Tavallaee M., Bagheri E., Lu W., and Ghorbani A. A., "A detailed analysis of the kdd cup 99 data set," in *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, Jul 2009, p. 1–6.

[12] Wawrowski Ł., Michalak M., Białas A., Kurianowicz R., Sikora M., Uchroński, and Kajzer A., "Detecting anomalies and attacks in network traffic monitoring with classification methods and xai-based explainability," *Procedia Comput. Sci.*, vol. 192, no. C, pp. 2259–2268, jan 2021.

[13] Wawrowski Ł, Białas A., Kajzer A., Kozłowski A., Kurianowicz R., Sikora M., Szymańska-Kwiecień A, Uchroński M., M. Białczak, Olejnik M., and Michalak M., "Anomaly detection module for network traffic monitoring in public institutions," *Sensors*, vol. 23, no. 6, 2023.

[14] Moustafa N., Ahmed M., and Ahmed S., "Data analytics-enabled intrusion detection: Evaluations of ton_iot linux datasets," in *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2020, pp. 727–735.

[15] Sperotto A., Sadre R., van Vliet F., and Pras A., "A labeled data set for flow-based intrusion detection," in *IP Operations and Management*, Giorgio Nunzi et al., Eds., Netherlands, Oct. 2009, Lecture Notes in Computer Science, pp. 39–50, Springer, 9th IEEE International Workshop on IP Operations and Management, IPOM 2009, IPOM ; Conference date: 29-10-2009 Through 30-10-2009.

[16] Moustafa N. and Slay J., "Unsw-nb15: a comprehensive data set for network intrusion detection systems (unsw-nb15 network data set)," in *2015 Military Communications and Information Systems Conference (MilCIS)*, 2015, pp. 1–6.

[17] Yang S., Kurose J., and Levine B., "Disambiguation of residential wired and wireless access in a forensic setting," in *2013 Proceedings IEEE INFOCOM*, 04 2013, pp. 360–364.

[18] Maede Zolanvari, "Wustl-iiot-2021 dataset for iiot cybersecurity research," https://www.cse.wustl.edu/~jain/iiot2/index.html, 2021.

[19] Ferrag M.A., Friha O., Hamouda D., Maglaras L., and Janicke H., "Edge-iiotset: A new comprehensive realistic cyber security dataset of iot and iiot applications: Centralized and federated learning," https://dx.doi.org/10.21227/mbc1-1h68, 2022.

[20] Garcia S., Parmisano A., and Erquiaga M. J., "IoT-23: A labeled dataset with malicious and benign IoT network traffic," Jan. 2020.

[21] Janusz A. and Ślęzak D., "KnowledgePit Meets BrightBox: A Step Toward Insightful Investigation of the Results of Data Science Competitions," in *Proceedings of the 17th Conference on Computer Science and Intelligence Systems, FedCSIS 2022, Sofia, Bulgaria, September 4-7, 2022*, 2022, vol. 30 of *Annals of Computer Science and Information Systems*, pp. 393–398.

[22] Andrzej Janusz, Daniel Kałuża, Agnieszka Chądzyńska-Krasowska, Bartek Konarski, Joel Holland, and Dominik Ślęzak, "IEEE BigData 2019 Cup: Suspicious Network Event Recognition," in *2019 IEEE International Conference on Big Data (IEEE BigData), Los Angeles, CA, USA, December 9-12, 2019*. 2019, pp. 5881–5887, IEEE.

[23] Andrzej Janusz, Mateusz Przyborowski, Piotr Biczyk, and Dominik Ślęzak, "Network Device Workload Prediction: A Data Mining Challenge at Knowledge Pit," in *Proceedings of the 2020 Federated Conference on Computer Science and Information Systems, FedCSIS 2020, Sofia, Bulgaria, September 6-9, 2020*, 2020, vol. 21 of *Annals of Computer Science and Information Systems*, pp. 77–80.

[24] Andrzej Janusz, Marek Sikora, Łukasz Wróbel, Sebastian Stawicki, Marek Grzegorowski, Piotr Wojtas, and Dominik Ślęzak, "Mining data from coal mines: Ijcrs'15 data challenge," in *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing: 15th International Conference, RSFDGrC 2015, Tianjin, China, November 20-23, 2015, Proceedings*. Springer, 2015, pp. 429–438.

[25] Adamczyk B., Brzęczek M., Michalak M., Kostorz I., WawrowskiŁ., Hermansa M., Czerwiński M., and Jamiołkowski A., "Dataset generation framework for evaluation of iot linux host–based intrusion detection systems," in *2022 IEEE International Conference on Big Data (Big Data)*, 2022, pp. 6179–6187.

[26] "MITRE ATT&CK® Adversarial Tactics, Techniques, and Common Knowledge," https://attack.mitre.org/versions/v13/, 2023.