

bradscholars

Machine Learning for Botnet Detection: An Optimized Feature Selection Approach

Item Type	Conference paper
Authors	Lefoane, Moemedi;Ghafir, Ibrahim;Kabir, Sohag;Awan, Irfan U.
Citation	Lefoane M, Ghafir I, Kabir S and Awan IU (2021) Machine Learning for Botnet Detection: An Optimized Feature Selection Approach. The 5th International Conference on Future Networks & Distributed Systems (ICFNDS 2021). December 15–16, 2021. Dubai, United Arab Emirates. ACM, New York. 6 pages.
DOI	https://doi.org/10.1145/3508072.3508102
Rights	© 2021 Association for Computing Machinery. Reproduced in accordance with the publisher's self-archiving policy. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org .
Download date	2026-03-05 23:13:24
Link to Item	http://hdl.handle.net/10454/18862

Machine Learning for Botnet Detection: An Optimized Feature Selection Approach

Moemedi Lefoane
University of Bradford
Bradford, United Kingdom
m.lefoane@bradford.ac.uk

Sohag Kabir
University of Bradford
Bradford, United Kingdom
s.kabir2@bradford.ac.uk

Ibrahim Ghafir
University of Bradford
Bradford, United Kingdom
i.ghafir@bradford.ac.uk

Irfan-Ullah Awan
University of Bradford
Bradford, United Kingdom
i.u.awan@bradford.ac.uk

ABSTRACT

Technological advancements have been evolving for so long, particularly Internet of Things (IoT) technology that has seen an increase in the number of connected devices surpass non IoT connections. It has unlocked a lot of potential across different organisational settings from healthcare, transportation, smart cities etc. Unfortunately, these advancements also mean that cybercriminals are constantly seeking new ways of exploiting vulnerabilities for malicious and illegal activities. IoT is a technology that presents a golden opportunity for botnet attacks that take advantage of a large number of IoT devices and use them to launch more powerful and sophisticated attacks such as Distributed Denial of Service (DDoS) attacks. This calls for more research geared towards the detection and mitigation of botnet attacks in IoT systems. This paper proposes a feature selection approach that identifies and removes less influential features as part of botnet attack detection method. The feature selection is based on the frequency of occurrence of the value counts in each of the features with respect to total instances. The effectiveness of the proposed approach is tested and evaluated on a standard IoT dataset. The results reveal that the proposed feature selection approach has improved the performance of the botnet attack detection method, in terms of True Positive Rate (TPR) and False Positive Rate (FPR). The proposed methodology provides 100% TPR, 0% FPR and 99.9976% F-score.

CCS CONCEPTS

• Security and privacy → Intrusion detection systems.

KEYWORDS

Botnet Detection, Machine Learning, Cyber Attacks, Internet of Things, Network security

ACM Reference Format:

Moemedi Lefoane, Ibrahim Ghafir, Sohag Kabir, and Irfan-Ullah Awan. 2021. Machine Learning for Botnet Detection: An Optimized Feature Selection Approach. In *The 5th International Conference on Future Networks & Distributed Systems (ICFNDS 2021)*, December 15–16, 2021, Dubai, United Arab Emirates. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3508072.3508102>

1 INTRODUCTION

Cybersecurity is a constantly evolving field. As such cybercriminals are constantly exploring and coming up with novel ways of finding vulnerabilities as well as exploiting them for malicious and illegal activities. Several attacks have been existing for years such as malware spread, the malwares are then used later to perform attacks such as data ex-filtration and Denial of Service attacks using or on infected devices [10, 11]. Covid-19 pandemic has led to increased remote services and remote working, rendering online security difficult and a challenging task while creating a golden opportunity for cybercriminals. This has resulted in an increased number of vulnerable devices leading to several attacks targeting the general public (online shopping fraud) and businesses. According to the Royal United Services Institute, these attacks have reached epidemic levels during covid-19 pandemic [35]. Evidently, Covid-19 has become one of the enabling factors for cyber attacks.

Increasingly consumer daily activities and industry operations heavily rely on IoT technology, with vast amounts of data collected from the environment and sent to the cloud for further analysis. Examples of such IoT solutions include the deployment of smart street lights, smart meters and air quality monitoring in smart cities, these are widely used in developed countries. The benefits of IoT systems range from conservation of energy and improved decision making as well addressing environmental concerns. While there are countless benefits brought by IoT technologies, it comes with a lot of security challenges. Some of the security issues are due to inherent limitations, the limitations include processing power and storage constraints [5, 25]. On the other hand, lack of sufficient configuration of IoT devices creates many weak links, therefore compromising IoT devices is usually not hard. Breaching IoT devices lead to more catastrophic damages such as taking out critical applications for an extended period, inevitably this results in loss of billions of dollars and, in the case of the healthcare sector, these attacks can lead to loss of lives [13, 24].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ICFNDS 2021, December 15–16, 2021, Dubai, United Arab Emirates

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8734-7/21/12...\$15.00

<https://doi.org/10.1145/3508072.3508102>

According to IoT analytics, the number of connected IoT devices reached 12 billion and surpassed the number of connected non IoT devices in 2020 [19]. The number of connected IoT devices continues to grow and is projected to reach more than 27 billion IoT connections by 2025 [26]. The proliferation of the IoT technologies across industries has opened up and increased attack surface/threat landscape, thus adding to the enabling factors for cyber attacks and acting more as a catalyst on some of the attacks like DDoS [18].

An opportunity of a lifetime is presented to cybercriminals who target these devices and infect them with bot malware, effectively turning them into botnets which are then used to carry out even more lethal and illegal attacks at a later stage. The term bot is short for robot, this is a type of malware that is usually transmitted by exploiting vulnerabilities of an application or operating system. Once a computer or device is infected by bot malware it becomes part of a botnet (robot network), the device is then automatically controlled by a command and control (C&C) server, as such bot infected devices would typically connect to C&C server and await instructions from the server. The C&C server is in-turn controlled by a bot master. Botnets are more powerful and sophisticated ammunition at the disposal of cybercriminals. Once a botnet is formed with a large number of infected devices, it is then used for different types of cyber attacks including renting for illegal activities such as crypto mining, data ex-filtration, phishing scams and DDoS [6, 12]. Indeed, in recent years botnets have been responsible for attacks such as DDoS, data ex-filtration etc. with the most common attacks being DDoS. Examples of such botnets include Mirai botnets and variations [21]. The botnets continue to evolve and in some instances new ones are derived from old ones. Web applications such as looking glass threat map show live attacks across the world, also indicating the infections per second [28, 33]. As indicated earlier, botnet attacks are a serious problem and result in the loss of billions for companies every year, therefore botnet detection and mitigation is an important problem that calls for more research from the cybersecurity research community. Several promising solutions proposed in the literature for botnet detection and mitigation include machine learning approaches.

This work proposes a machine learning approach to botnet detection and mitigation by analysing network traffic derived dataset. The focus of the proposed approach is to reduce dimensions during feature selection as part of a machine learning solution. The feature selection identifies noisy features based on the frequency of occurrence of value counts for each of the features. Once identified the noisy features are removed.

The rest of the paper is organised as follows: Section 2 presents the related work, Section 3 provides details of the proposed approach. Section 4 discusses the experimental setup, dataset used and evaluation of results. Finally, Section 5 concludes the paper.

2 RELATED WORK

Detection of botnet activities on network traffic is a challenging task as botnets keep evolving. Several techniques have been proposed in the literature ranging from signature-based approaches to machine learning based anomaly detection approaches [9]. Signature-based detection systems are typically deployed on network intrusion detection systems such as Zeek [2] and Snort [1, 7, 15]. They yield

good performance for attacks with known signatures, however, detecting zero day attacks is still a challenge. The proliferation of IoT solutions across industries exacerbates botnet detection challenges, particularly because IoT devices are easy to compromise and deploy in large numbers.

Several machine learning approaches have been investigated for the detection of botnet attacks. Specifically, both supervised and unsupervised learning. The popular machine learning algorithms deployed for malicious activities detection are support vector machines, random forest, logistic regression and decision trees [4, 14, 16, 17, 31, 32, 34]. In other studies, algorithms deployed for detection of botnet attacks include artificial neural networks deployed together with machine learning [27]. Deep learning is also proposed as one of the best models to deploy for botnet detection but not suitable for deployment in IoT devices as they are typically memory constrained, for those reasons solutions such as reducing dimension of features is often proposed [23]. Joshi and Abdelfattah [17] investigated the efficacy of various machine learning algorithms on IoT botnet attack classification. Wiyono et al [34] analysed the performance of the decision tree deployed as classification model for botnet activities in IoT network forensics.

Intrusion Detection Systems (IDSs) are also promising solutions for botnet detection and have been proposed in the literature [3]. IDS is an active field of research, particularly when incorporating anomaly detection techniques, and are even better when deployed for detection of malicious activities in real time. Other works include detection of the lifecycle of attacks such as botnet attacks. Here, correlation approaches are proposed to help reduce false alarms, leading to improved performance [30]. Other approaches include analysis aimed at understanding how botnets spread and have the potential to learn how botnets evolve and could lead to improved and effective techniques that can be used to detect botnet attacks. In their paper, Mahboubi et al. [20] investigated and explored the use of epidemiological modelling to help understand how botnets spread. The reported evaluation yielded promising results that have the potential to improve detection approaches by deriving new signatures from the understanding of how botnets evolve and spread over time. Other works focus on specific phases of a botnet attack, for example detecting C&C communication from the servers, which in turn aims to prevent further attacks by detecting early phases of a botnet attack [37].

Blockchain technology is an emerging and interesting technology applied to a number of areas. It is a technology that is explored for addressing data integrity, privacy, and authorization issues in IoT [22, 38]. Spathoulas et al. [29] proposed a blockchain based detection of DDoS attacks from IoT botnets, the proposed approach is an agent-based detection system deployed on the gateway instead of IoT devices.

Velasco-Mata et al. [32] proposed botnet attacks detection approach, with feature selection based on information gain and gini importance, the results of the work show improved performance. Likewise, Injadat et al. [16] proposed an optimised machine learning approach for detection of botnet attacks and reported improved performance. While statistical feature selection and reduction techniques such as chi square already exist in the literature, several alternative feature selection approaches are often proposed and deployed together with machine learning resulting in improved

performance, as such it is one aspect of detection approach that is worth exploring when it comes to boosting performance of existing machine learning approaches. This work investigates the usefulness of features and proposes a feature selection approach that identifies and removes less influential features from the network traffic data. The effect of the proposed feature selection approach is evaluated on three machine learning classification algorithms. The proposed approach is investigated in the context of IoT, as such IoT traffic generated data is used to evaluate the proposed method. The proposed approach focuses on extracting the usefulness of a feature based on the most dominant pattern of each of the features, the frequency of occurrence of the most dominant pattern is used as a score, finally, a manually selected threshold is then used to identify and remove noisy features with score above threshold value. To the best of our knowledge addressing feature selection this way has not been explored in the literature.

3 PROPOSED METHODOLOGY

The proposed methodology aims to classify the network traffic into malicious and normal, for the purpose of detecting botnet traffic. This work investigates the effect of eliminating noisy features and keeping only the most influential features on binary classification of normal and malicious instances. The methodology is set up in two stages; the first one is feature selection and the second stage is building the detection model. Figure 1 illustrates the proposed methodology. In the first stage data is cleaned and noisy features are detected and removed from the data, then three algorithms are trained and evaluated to build the detection model. Subsection 3.1 and 3.2 elaborate more on the two stages.

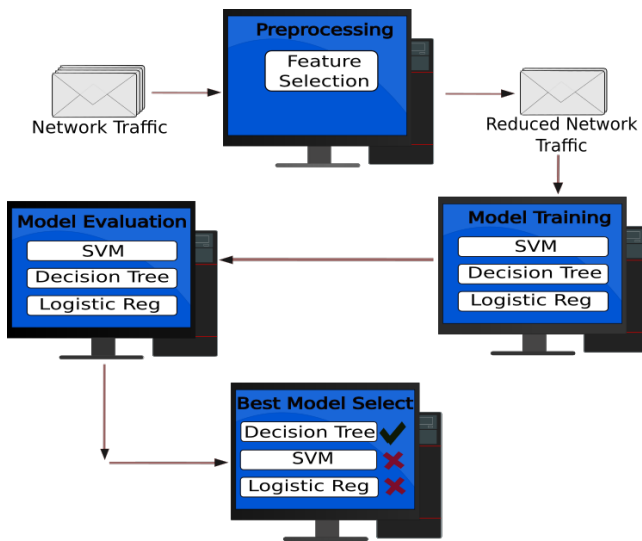


Figure 1: Proposed Methodology for the detection of Botnet traffic.

3.1 Feature Selection

The main focus for this stage is the identification of features from the data that are less influential for classification purposes. Once

identified, these features are then removed from the data before proceeding to the next stage. The proposed approach focuses on the most dominant pattern for each of the features. A frequency of occurrence for each of the dominant pattern values with respect to the total instances is determined and analysed. Ordering the frequencies of the dominant patterns in descending order, it is observed that the frequencies of the most dominant patterns range from 100% down to the lowest scoring frequencies. A further analysis of the samples for each of the feature frequencies, reveals that features with higher frequencies tend to have instances from all the target classes and those in with lower frequencies have only one target class, thus lower frequency pattern features are more influential for classification purposes than ones with higher frequencies which are noisy features, with level of noise increasing as frequency increases. The proposed approach therefore targets these noisy features occurring at high frequencies, with a manually selected threshold value set as a cut-off for removal of the noisy features.

Algorithm 1 provides the steps followed for identification and removal of less informative features. The algorithm requires source port and target column names, also threshold value is required in the form of percentage for example 100.0. The reason for requiring positions of obvious features is that in network traffic data, features such as source port number are not useful as they are generated randomly as such would not have any meaningful contribution on classification. Furthermore, a unique identifier is another obvious one that does not contribute to detection of malicious activities in network traffic data. In line 1 and 2 total observations in data is computed and saved in r and an empty list (col_dl) to store names of less influential features is created. Source port column names form part of the obvious features that are less useful and the target name is identified here as only features are to be analysed at this stage. The given features are removed as indicated in line 3 – 5. The threshold parameter should be between 0.0 and 1.0. Line 6 – 11 identifies column names that are less influential based on the threshold provided, this is done by iterating through each column, unique value counts are then computed and if any unique count is greater than threshold then the column name is added to less influential features list. Finally, all the columns identified as less informative are removed from the data. At this point the data is ready for stage 2.

3.2 Model Training and Best Model Selector

This stage takes as input reduced data from the first stage and is concerned with processing the data further before training several classification algorithms. As the goal of the proposed method is a binary classification of malicious and normal, all the malicious classes are combined into one class. The resulting target labels are only normal and malicious. Once further processing is done, the traffic data is passed into model training, this is where three classification algorithms are trained, specifically decision tree, support vector machine and logistic regression. The choice of these algorithms is based on their popularity in the literature [4, 17, 31, 32] that address similar research problems. The trained models are then passed to model evaluation, which evaluates and produces performance results. The evaluation is performed with 5 fold cross

Algorithm 1 Implementation Pseudo-code for Selection of Most Influential features

```

Require: data                                ▶ Dataset
Require: sp                                  ▶ Source port provided by user
Require: tgt                                  ▶ target provided by user
Require: thld                                ▶ frequency of feature value
1:  $r \leftarrow \text{compute total instances (rows)}$ 
2:  $col\_dl[] \leftarrow ""$  ▶ Create empty list for feature names to drop
3: drop sp
4: drop tgt
5: drop id_col
6: for col_name in columns do
7:    $v[] \leftarrow \text{compute total count for each unique value}$ 
8:   for uniq_value in col_name do
9:     if  $v[uniq\_value]/r > thld$  then
10:      if col_name not in col_dl then
11:         $col\_dl[] \leftarrow \text{Add col\_name to drop list}$ 
12:      end if
13:    end if
14:  end for
15: end for

```

validation [36]. The results for each model go through the best model selector which analyses performance metrics for each of the machine learning algorithms trained. Finally, based on the results of the trained models, the best performing model is selected based on the overall score for each classification algorithm on all chosen metrics.

4 EVALUATION OF RESULTS

The focus of the experiments is the classification of normal and malicious connections from the network traffic data, the malicious labels for the dataset namely: C&C, DDoS and PartOfHorizontalScan were combined into one class. This resulted in two classes: normal and malicious. To evaluate the effectiveness of the proposed method, a publicly available dataset IoT23 [8] is used. Experiments are performed in two scenarios. In the first scenario, experiment is performed without the feature selection stage, this is done as a reference to determine if the proposed method given in Section 3 improves the performance. In the second scenario, the experiment is performed after removing obvious less useful features such as unique ID, time stamp, source IP address, source port number. Sections 4.1 and 4.2 provide more details on the experimental setup, which includes the results of the experiments.

4.1 Dataset Description

IoT23 is a dataset derived from IoT devices network traffic. Refer to Table 1 for label distribution of the chosen dataset. This dataset is labelled connection logs files generated by Zeek from the network traffic. The dataset has 20 feature variables and 4 classes in the target variable indicated in Table 1.

Table 1: CTU-IoT-Malware-Capture-34-1 (Mirai) Labels Distribution

Label	Flows
Normal	1923
C & C	6706
DDoS	14394
PartOfHorizontalPortScan	122

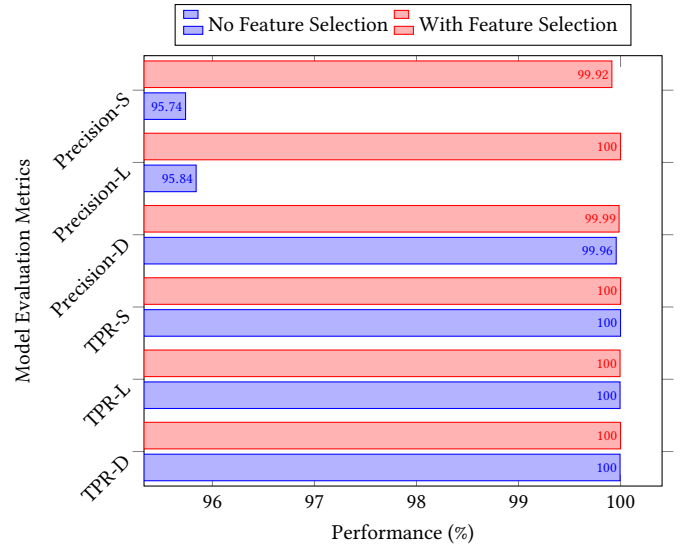


Figure 2: True Positive Rate and Precision Evaluation Results for Several Classification Model

4.2 Experimental Setup

Following removal of the obvious less useful features already stated, the results here are 16 features data. The steps outlined in Section 3.2 were deployed on the data to generate results for the first scenario.

The second experiment was performed following the approach outlined in Section 3, specifically the steps outlined in Section 3.1 and Section 3.2 were deployed. The results of the first stage, outlined in 3.1 are six features only data, compared to first scenario this shows that ten less informative features were identified by proposed method and removed.

The evaluation metrics computed to evaluate the performance of the proposed methodology are: True Positive Rate (*TPR*), a proportion of correctly classified malicious connections to total malicious instances, this is effectively the detection rate that measures how accurately a model predict malicious connections in the data. False Positive Rate (*FPR*) is a measure of proportion of false alarms to total normal connections, *Precision* is a proportion of correctly classified malicious connections to a total classified as malicious, this measure finds out how many of the connections predicted as malicious are actually malicious. Overall Success Rate (*OSR*) is a proportion of all correctly classified to total instances and *F-score* is a measure of success based on trade off between *Precision* and *TPR*.

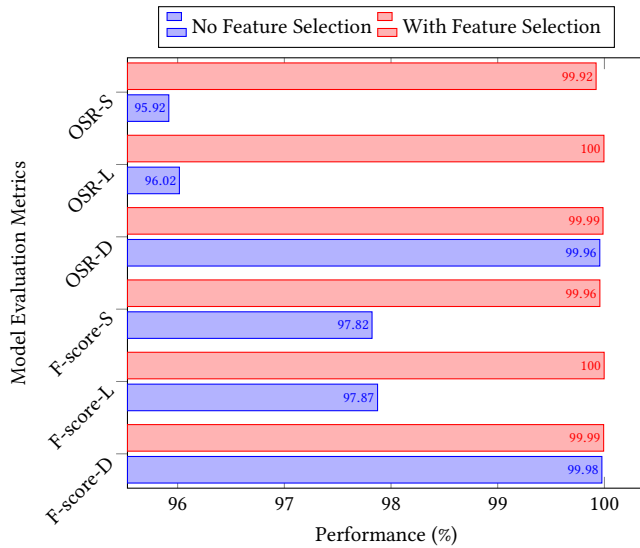


Figure 3: F-score and OSR Evaluation Results for Several Classification Model

Table 2, Figure 2 and 3 shows *FPR*, *TPR*, *Precision*, *OSR* and *F – score* performance results for the experiments performed. *NFS*, No Feature Selection is applied, refers to the experimental scenario where only part of the proposed method explained in Section 3.2 is applied. *WFS*, With Feature Selection, is a scenario where the proposed method is applied fully as detailed in Section 3.1 and 3.2. *FPR – D*, *FPR – L* and *FPR – S* indicate results for False Positive Rate for decision tree, logistic regression and support vector machine respectively. *Precision – D*, *Precision – L* and *Precision – S* refers to Precision results for decision tree, logistic regression and support vector machines respectively. Finally *OSR – D*, *OSR – L* and *OSR – S* refers to overall success rate results for decision tree, logistic regression and support vector machines respectively. The results reveal that *TRP*, *FPR*, *Precision*, *F – score* and *OSR* on the proposed approach consistently improved the performance results and overall decision tree performed across all the metrics evaluated, therefore decision tree (Fine Tree) was selected as the best classifier of normal and malicious connections in network traffic data.

Table 2: False Positive Rate for classification Models

Experiment	FPR-D	FPR-L	FPR-S
NFS	0.468	47.894	49.142
WFS	0.156	0	0.936

5 CONCLUSION

This paper presented an optimized feature selection approach for botnet detection. The feature selection is based on the frequency of occurrence of the value counts in each of the features with respect to total instances. Then three machine learning algorithms (decision tree, logistic regression, and support vector machine) are

explored to build the best detection model, utilizing the reduced network traffic. The results show that the proposed method consistently improves the performance across all the measures: *TPR*, *FPR*, *Precision*, *F – score* and *OSR*. When models for the detection of malicious activities are deployed, the goal is to train the model such that it minimises false alarms while also increases the detection rate, the proposed approach consistently improve performance in that regard.

REFERENCES

- [1] 2021. Snort. <https://www.snort.org/>. Accessed: 2021-10-20.
- [2] 2021. Zeek. <https://zeek.org/>. Accessed: 2021-10-20.
- [3] Eirini Anthi, Lowri Williams, Malgorzata Slowinska, George Theodorakopoulos, and Pete Burnap. 2019. A Supervised Intrusion Detection System for Smart Home IoT Devices. *IEEE Internet of Things Journal* 6, 5 (2019), 9042–9053. <https://doi.org/10.1109/JIOT.2019.2926365>
- [4] Rohan Bapat, Abhijith Mandya, Xinyang Liu, Brendan Abraham, Donald E. Brown, Hyoungjung Kang, and Malathi Veeraraghavan. 2018. Identifying malicious botnet traffic using logistic regression. In *2018 Systems and Information Engineering Design Symposium (SIEDS)*. 266–271. <https://doi.org/10.1109/SIEDS.2018.8374749>
- [5] Sana Belguith, Nesrine Kaaniche, Mohammad Hammoudeh, and Tooska Dargahi. 2020. Proud: Verifiable privacy-preserving outsourced attribute based signcryption supporting access policy update for cloud assisted iot applications. *Future Generation Computer Systems* 111 (2020), 899–918.
- [6] D.M. Diab, B. AsSadhan, H. Binsalleeh, S. Lambbotharan, K.G. Kyriakopoulos, and I. Ghafir. 2021. Denial of service detection using dynamic time warping. *International Journal of Network Management* (2021). <http://hdl.handle.net/10454/18458>
- [7] RaviTeja Gaddam and M. Nandhini. 2017. An analysis of various snort based techniques to detect and prevent intrusions in networks proposal with code refactoring snort tool in Kali Linux environment. In *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*. 10–15. <https://doi.org/10.1109/ICICCT.2017.7975177>
- [8] Sebastian Garcia, Agustin Parmisano, and Maria Jose Erquiaga. 2020. *IoT-23: A labeled dataset with malicious and benign IoT network traffic*. <https://doi.org/10.5281/zenodo.4743746> More details here <https://www.stratosphereips.org/datasets-iot23>.
- [9] Ibrahim Ghafir, Martin Husak, and Vaclav Prenosil. 2014. A survey on intrusion detection and prevention systems. In *Proceedings of student conference Zvule, IEEE/UREL. Brno University of Technology*, Vol. 1014.
- [10] Ibrahim Ghafir, Konstantinos G. Kyriakopoulos, Francisco J. Aparicio-Navarro, Sangarapillai Lambbotharan, Basil Assadhan, and Hamad Binsalleeh. 2018. A Basic Probability Assignment Methodology for Unsupervised Wireless Intrusion Detection. *IEEE Access* 6 (2018), 40008–40023. <https://doi.org/10.1109/ACCESS.2018.2855078>
- [11] Ibrahim Ghafir, Konstantinos G. Kyriakopoulos, Sangarapillai Lambbotharan, Francisco J. Aparicio-Navarro, Basil Assadhan, Hamad Binsalleeh, and Diab M. Diab. 2019. Hidden Markov Models and Alert Correlations for the Prediction of Advanced Persistent Threats. *IEEE Access* 7 (2019), 99508–99520. <https://doi.org/10.1109/ACCESS.2019.2930200>
- [12] Ibrahim Ghafir, Vaclav Prenosil, Mohammad Hammoudeh, Francisco J. Aparicio-Navarro, Khaled Rabie, and Ahmad Jabban. 2018. Disguised Executable Files in Spear-Phishing Emails: Detecting the Point of Entry in Advanced Persistent Threat. In *Proceedings of the 2nd International Conference on Future Networks and Distributed Systems (Amman, Jordan) (ICFNDS '18)*. Association for Computing Machinery, New York, NY, USA, Article 44, 5 pages. <https://doi.org/10.1145/3231053.3231097>
- [13] Mohammad Hammoudeh, Ibrahim Ghafir, Aheçene Bounceur, and Thomas Rawlinson. 2019. Continuous Monitoring in Mission-Critical Applications Using the Internet of Things and Blockchain. In *Proceedings of the 3rd International Conference on Future Networks and Distributed Systems (Paris, France) (ICFNDS '19)*. Association for Computing Machinery, New York, NY, USA, Article 27, 5 pages. <https://doi.org/10.1145/3341325.3342018>
- [14] Mandira Hegde, Gilles Kepnang, Mashail Al Mazroei, Jeffrey S. Chavis, and Lanier Watkins. 2020. Identification of Botnet Activity in IoT Network Traffic Using Machine Learning. In *2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*. 21–27. <https://doi.org/10.1109/IDSTA50958.2020.9264143>
- [15] Xiaojin Hong, Changzhen Hu, Zhigang Wang, Guoqiang Wang, and Ying Wan. 2012. VisSRA: Visualizing Snort Rules and Alerts. In *2012 Fourth International Conference on Computational Intelligence and Communication Networks*. 441–444. <https://doi.org/10.1109/CICN.2012.207>

- [16] MohammadNoor Injadat, Abdallah Moubayed, and Abdallah Shami. 2020. Detecting Botnet Attacks in IoT Environments: An Optimized Machine Learning Approach. In *2020 32nd International Conference on Microelectronics (ICM)*. 1–4. <https://doi.org/10.1109/ICM50269.2020.9331794>
- [17] Shreehar Joshi and Eman Abdelfattah. 2020. Efficiency of Different Machine Learning Algorithms on the Multivariate Classification of IoT Botnet Attacks. In *2020 11th IEEE Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*. 0517–0521. <https://doi.org/10.1109/UEMCON51285.2020.9298095>
- [18] Georgios Kambourakis, Constantinos Kolia, and Angelos Stavrou. 2017. The Mirai botnet and the IoT Zombie Armies. In *MILCOM 2017 - 2017 IEEE Military Communications Conference (MILCOM)*. 267–272. <https://doi.org/10.1109/MILCOM.2017.8170867>
- [19] Knud Lasse Lueth. 2021. State of the IoT 2020: 12 billion IoT connections, surpassing non-IoT for the first time. <https://iot-analytics.com/state-of-the-iot-2020-12-billion-iot-connections-surpassing-non-iot-for-the-first-time/>. Accessed: 2021-10-07.
- [20] Arash Mahboubi, Seyit Camtepe, and Keyvan Ansari. 2020. Stochastic Modeling of IoT Botnet Spread: A Short Survey on Mobile Malware Spread Modeling. *IEEE Access* 8 (2020), 228818–228830. <https://doi.org/10.1109/ACCESS.2020.3044277>
- [21] Joel Margolis, Tae Tom Oh, Suyash Jadhav, Young Ho Kim, and Jeong Noyo Kim. 2017. An In-Depth Analysis of the Mirai Botnet. In *2017 International Conference on Software Security and Assurance (ICSSA)*. 6–12. <https://doi.org/10.1109/ICSSA.2017.12>
- [22] Bhabendu Kumar Mohanta, Debasish Jena, Somula Ramasubbarreddy, Mahmoud Daneshmand, and Amir H. Gandomi. 2021. Addressing Security and Privacy Issues of IoT Using Blockchain Technology. *IEEE Internet of Things Journal* 8, 2 (2021), 881–888. <https://doi.org/10.1109/JIOT.2020.3008906>
- [23] Segun I. Popoola, Bamidele Adebisi, Mohammad Hammoudeh, Guan Gui, and Haris Gacanim. 2021. Hybrid Deep Learning for Botnet Attack Detection in the Internet-of-Things Networks. *IEEE Internet of Things Journal* 8, 6 (2021), 4944–4956. <https://doi.org/10.1109/JIOT.2020.3034156>
- [24] Umar Raza, James Lomax, Ibrahim Ghafir, Rupak Kharel, and Ben Whiteside. 2017. An IoT and Business Processes Based Approach for the Monitoring and Control of High Value-Added Manufacturing Processes. In *Proceedings of the International Conference on Future Networks and Distributed Systems (Cambridge, United Kingdom) (ICFNDS '17)*. Association for Computing Machinery, New York, NY, USA, Article 37, 8 pages. <https://doi.org/10.1145/3102304.3102341>
- [25] Jibrán Saleem, Mohammad Hammoudeh, Umar Raza, Bamidele Adebisi, and Ruth Ande. 2018. IoT standardisation: Challenges, perspectives and solution. In *Proceedings of the 2nd international conference on future networks and distributed systems*. 1–9.
- [26] Satyajit Sinha. 2021. State of IoT 2021: Number of connected IoT devices growing 9% to 12.3 billion globally, cellular IoT now surpassing 2 billion. <https://iot-analytics.com/number-connected-iot-devices/>. Accessed: 2021-10-07.
- [27] Yan Naung Soe, Yaokai Feng, Paulus Insap Santosa, Rudy Hartanto, and Kouichi Sakurai. 2020. Machine Learning-Based IoT-Botnet Attack Detection with Sequential Architecture. *Sensors* 20, 16 (2020). <https://doi.org/10.3390/s20164372>
- [28] Looking Glass Cyber Solutions. 2019. Looking Glass Threat Map. <https://map.lookingglasscyber.com/>. Accessed: 2021-10-07.
- [29] Georgios Spathoulas, Nikolaos Giachoudis, Georgios-Paraskevas Damiris, and Georgios Theodoridis. 2019. Collaborative Blockchain-Based Detection of Distributed Denial of Service Attacks Based on Internet of Things Botnets. *Future Internet* 11, 11 (2019). <https://doi.org/10.3390/fi11110226>
- [30] Kalupahana Liyanage Kushan Sudheera, Dinil Mon Divakaran, Rhishi Pratap Singh, and Mohan Gurusamy. 2021. ADEPT: Detection and Identification of Correlated Attack Stages in IoT Networks. *IEEE Internet of Things Journal* 8, 8 (2021), 6591–6607. <https://doi.org/10.1109/JIOT.2021.3055937>
- [31] Li Suhuan and Huang Xiaojun. 2019. Android Malware Detection Based on Logistic Regression and XGBoost. In *2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS)*. 528–532. <https://doi.org/10.1109/ICSESS47205.2019.9040851>
- [32] Javier Velasco-Mata, Víctor González-Castro, Eduardo Fidalgo Fernández, and Enrique Alegre. 2021. Efficient Detection of Botnet Traffic by Features Selection and Decision Trees. *IEEE Access* 9 (2021), 120567–120579. <https://doi.org/10.1109/ACCESS.2021.3108222>
- [33] Steven Walker-Roberts, Mohammad Hammoudeh, Omar Aldabbas, Mehmet Aydin, and Ali Dehghantanha. 2020. Threats on the horizon: Understanding security threats in the era of cyber-physical systems. *The Journal of Supercomputing* 76, 4 (2020), 2643–2664.
- [34] Rizky Tri Wiyono and Niken Dwi Wahyu Cahyani. 2020. Performance Analysis of Decision Tree C4.5 as a Classification Technique to Conduct Network Forensics for Botnet Activities in Internet of Things. In *2020 International Conference on Data Science and Its Applications (ICoDSA)*. 1–5. <https://doi.org/10.1109/ICoDSA50139.2020.9212932>
- [35] Helena Wood, Tom Keatinge, Keith Ditcham, and Ardi Janjeva. 2021. The Silent Threat: The Impact of Fraud on UK National Security. <https://rusi.org/explore-our-research/publications/occasional-papers/silent-threat-impact-fraud-uk-national-security>. Accessed: 2021-10-07.
- [36] Sanjay Yadav and Sanyam Shukla. 2016. Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. In *2016 IEEE 6th International Conference on Advanced Computing (IACC)*. 78–83. <https://doi.org/10.1109/IACC.2016.25>
- [37] Lihua Yin, Xi Luo, Chunsheng Zhu, Liming Wang, Zhen Xu, and Hui Lu. 2020. ConnSpoller: Disrupting C and C Communication of IoT-Based Botnet Through Fast Detection of Anomalous Domain Queries. *IEEE Transactions on Industrial Informatics* 16, 2 (2020), 1373–1384. <https://doi.org/10.1109/TII.2019.2940742>
- [38] Ma Zhaofeng, Wang Lingyun, Wang Xiaochang, Wang Zhen, and Zhao Weizhe. 2020. Blockchain-Enabled Decentralized Trust Management and Secure Usage Control of IoT Big Data. *IEEE Internet of Things Journal* 7, 5 (2020), 4000–4015. <https://doi.org/10.1109/JIOT.2019.2960526>