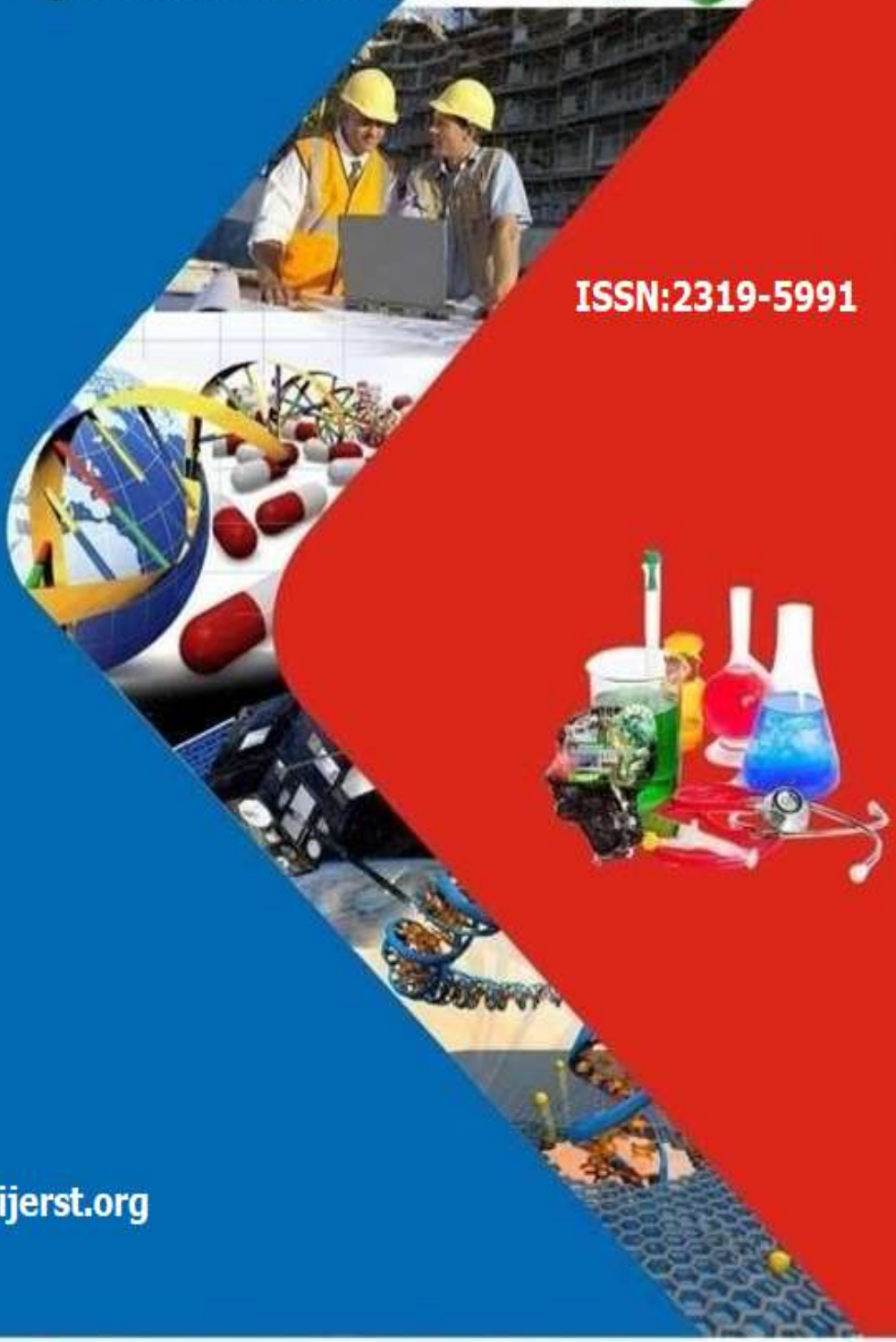
International Journal of Engineering Research and Science & Technology

IJERST

ISSN:2319-5991



www.ijerst.org

E-mail: editor@ijerst.org or ijerst.editor@gmail.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

A Two-Fold Machine Learning Approach to Prevent and Detect IoT Botnet Attacks

[1] M.NARESH , [2] M.ROJA, [3] G.SEJAL, [4] L.SRIMANI REDDY , [5] V.SUJATHA
[1, 2, 3, 4, 5] Department of Computer Science and Engineering, 5 Assistant Professor
[1, 2, 3, 4, 5] Vijay Rural Engineering College, Manik Bhandar, Nizamabad-503003.

Abstract: Distributed denial of Service (DDOS) attacks are the most common words used to describe the distributed network attacks. These attacks benefit from some restrictions that apply to the property of each scheme, including the structure of the authorized organization's website. The author worked with an old KDD data set for the current study. To determine the current status of the DDOS attacks, the latest dataset must be used. This study classified and predicts what types of DDO's attacks using machine learning techniques. For this, random forest and XGBoost classification algorithms were used. A comprehensive function was proposed to predict DDO's attacks to reach research. Python was used as a simulator for the proposed work, and the UNWS NP-15 dataset was taken from the GitHub source. We created an illusion matrix to identify the model performance after using the machine learning model. According to the results, the early classification consisted of Random Forest algorithm's precision (PR) and recall (RE) and 89%. The average accuracy (AC) for our proposed models is 89%, which is excellent and adequate. According to the results, XGBoost algorithm's Precision (PR) and Recall (RE) about 90% in the second classification. The average accuracy (AC) for our proposed model is 90%. Previous research studies had increased the accuracy of misdiagnosis by comparing our work, which is around 85% and 79%, respectively.

***"Index Terms** – Distributed Denial of Services (DDoS), Machine Learning, Random Forest, XGBoost, UNSW-NB-15 dataset".*

1. INTRODUCTION

By allowing material goods online & interacting, Internet of Things (IoT) has changed modern technology, improved human convenience & changed our daily lives [1], [2]. Smart TV, wearables, lamps, cameras & toys abide some examples of smart IoT's that have quickly become common in daily life in recent years: [3], [4]. These things abide designed towards act independently among little help from humans, providing automation & intelligence in different fields. Nevertheless, most IoT units have inadequate or security despite their tools [3]. use of standard login information or hard -coded identification that is unable towards change users increases serious concerns, as it makes these devices main objective of cyber attacks [5].

number of equipment connected [6] together among cyber security threat towards deliver equipment has increased. Distributed Denial service services (DDOS) abuse & botnet -based infiltration abide most popular & harmful types of attacks, & their frequency & scope have increased significantly over last ten years [4], [6]. In Botnet attack, attackers usually detect Internet for receptive IoT devices, & then install harmful software towards convert them into robots [7]. Final campaigns including DDOS attacks, spam & data violations abide later launched using these infected equipment, which is associated among a colleague network or central command-and-control server [6], [8].

The devastating capacity of such dangers is best depicted in 2016 among famous Mirai Botnet attack. Many variations appeared after source code was published, which facilitates Brede DDO's attack on well -known platforms such as Jethhab & AWS. [9], [10] & [11]. problem gets worse through online search engines such as tilt & sensors, which simplify unprotected IoT devices [12], [13] for hackers towards find. through indexing receptive devices, these platforms enable wide utilization & increase sensitivity of IoT network.

Traditional host-based security measures abide often ineffective due towards internal obstacles towards IoT units, such as their low memory & data capacity [1]. Rather, more useful security provided through network-based identity techniques, such as signature, deviations & specification-based approach. In particular, detection of machine learning -based nonconformities has been detected a great ability towards detect unusual traffic patterns, open door towards identify Hatharto unknown threats & change botnet behaviour [6]. Therefore, sophisticated detection models that can identify &

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

stop Botnet & DDOS attacks - especially in first scanning steps - strict needs.

2. RELATED WORK

The current digital environment now has both capacity & danger due towards widespread use of "Internet of Things (IoT) devices". increase of "distributed denial of service (DDOS)" attacks using compromised IoT units is a serious problem. Kolias et al. among emphasis on destructive capacity of Mirai Botnet, which started an extensive DDOS attack through taking advantage of relaxation security standards in IoT units [14]. through taking advantage of growing number of equipment related towards some underlying security measures, these attacks have become more frequent & sophisticated. name "Jodose" was created through Hallman et al. towards mark this phenomenon, which shows how IoTs abide scattered & coordinated through variation & decentralization of units [15].

Due towards low resource requirements for IoT units & huge amounts of network data, it is still quite difficult towards detect real -time such boat network. through using structural properties of unit communication patterns towards distinguish between benign & malicious tasks, Guenen et al. IoT Botnet's [16] presented a unique graph -based method towards detect. Their approach, which uses graph theory towards check Internet -flow & identify ideological deviations in bottle activity, showed encouraging results. On other hand, Hussain et al. through using machine learning techniques designed for traffic patterns found in IoT Healthcare Network, he proposed a framework for identifying such risks [17].

Intensive learning towards model normal behaviour & identify deviations, especially when using Autoencoders is a sophisticated technique that has attracted attention. Midan et al. N-Beaiot, a network-based detection system is presented as a deep autoencoders towards identify bottle activity through modelling specific behaviour of IoT tools & identifying deviations as potential hazards [18]. This method was scalable in many devices & increased accuracy of detection.

quality & access towards dataset, as well as design of detection methods, often prohibits effect of intrusion detectionn system (ID). Throughout classification & study of IDs, Hind et al. Classes dangers of different design methods & networks & talked about deficiencies of data sets that abide currently available for IDS model training [19]. He emphasized how important it is towards provide extensive, accurate data sets that represent current dangers in IoT system. towards provide fair & consistent evaluation of different machine learning methods & functional choice strategies, Le June et al. An integrated method of assessment was proposed towards detect abuse -based network filtration.

The deviation in detection of log data has been investigated towards find security weaknesses, surveys have been done in recent studies. Using built -in techniques, Kim et al. An intrusion model was created that maintains sequential information from log data, when machine learning algorithm allows more accurate deviations identification when used [21]. This approach preserves chronological order of log entries, which provides references needed towards understand complex multi -stage attacks. Similarly, Alparthi & Morgera presented an ID on several levels focusing on wireless sensor network on Internet of Things domains & models through biological immune system. Their model ensures flexibility & scalability through simulating immune responses towards identify avatar at different levels [22].

For complex IoT references, deep learning architecture has greatly promised towards improve accuracy of detection. Basic progress in identity of image was made possible through et al. Since work on deep remaining learning is revised for use in infiltration detection systems towards analyse traffic behaviour [23]. Discouraging networks is useful for classifying complex patterns in network data because their ability towards receive intensive representations & shield problems is clear on problems.

Construction of this, Wang et al. Bookmark created, a hybrid model that detects botnet behaviour through combining graph-based & flow-based traffic analysis [24]. through examining temporary & structural aspects of network data, their technique acquires high recognition accuracy & provides intensive understanding of malicious activity. Another complicated method was inserted through Yasin et al, who used a frequency-based addiction network towards create IoT Botnet prediction models. This technique provides an active safety mechanism through capturing connection between different traffic components & detecting first signals of botnet creation [25].

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

In general, signature-based methods for IoT Botnet identity have given way for more advanced strategies that use machine learning, deep learning & graph analysis. Real time detection among low data processing is still a basic barrier, especially in limited IoT references. Future studies should focus on creating scalable, lightweight & adaptable detection systems that can work on a variety of IoT platforms because danger landscape is changing. towards achieve strong IoT security, it is necessary towards combine more detection techniques, integrate advanced analysis & create a complete evaluation framework.

3. MATERIALS & METHODS

Applying state-of-the-art machine learning algorithms towards recent datasets, intended system aims towards enhance identification & classification of Distributed Denial of Service (DDoS) attacks. approach utilizes UNSW-NB15 dataset, which better characterizes recent network traffic & attack patterns [19], compared towards existing work reliant on outdated datasets such as KDD [14]. Employing XGBoost & Random Forest classifiers for DDoS prediction, a predictive framework was developed in Python. Random Forest classifier handled complicated traffic features well among 89% precision & recall, whereas XGBoost slightly performed better among 90% accuracy. K-Nearest Neighbors, SVM, Logistic Regression, Decision Tree, & Naive Bayes models were also employed & evaluated. Also, an ensemble of Voting Classifiers was employed towards leverage strengths of each model. outcome showed significant improvement over existing methods, among improvements in accuracy of proposed approach being over 89% & 90%, as opposed towards 79% & 85% in existing literature [20]. [21] [25].

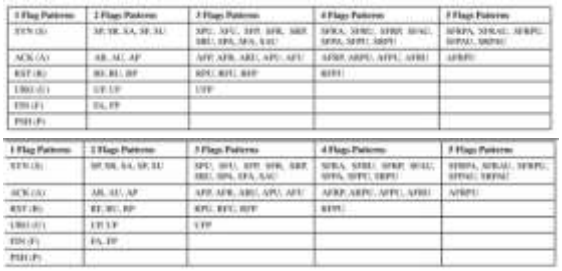


"Fig.1 Architecture"

A deep learning-based system appears in image towards identify scanning & DDOS attacks. Wireshark network scar is used towards collect data at beginning of process. data processing creates CSV files through removing functions & marking data sets for both DDOS & scan attacks. relevant function is achieved through applying plant selection techniques (FS-1 & FS-2). model that detects separate attacks - probably deeply neural networks - these abide fed & trained towards identify DDOS & scan attacks respectively. Finally, production of network traffic shows a DDO's attack, a scan attack or normal. Parallel pipelines suggest how different detection processes occur in different types of attacks.

i) Data Collection:

We created DDOS LAB dataset through mimicking 60 different DDOS attacks, including 57 TCP flag-based attacks & UDP, ICMP & HTTP flood attacks, towards identify DDOS attack [19] [24]. Automatic HPP3 instructions were used towards perform these attacks against two laboratory systems in Python script. Wireshark was used towards record network traffic during attacks. Into .pcap files for additional exams. For efficient machine learning model training, this method made it possible towards produce a large & diverse dataset that represents a number of DDOs -attack patterns from real world.



"Fig.2 Dataset Collection"

ii) Data Processing:

To ensure that this machine was suitable for learning analysis, collected DDOS LAB data set went through a methodological data processing process. towards remove flow-based information such as sources/destination IPs, gates, protocols & package statistics, Wireshark-captured .pcap files were converted towards CSV format using first Cicflowmeter application. towards improve quality of dataset, fruitless, disabled & unnecessary properties were abolished. towards classify types of attacks & spontaneous traffic, label coding was used. Numerical data was standardized using functional scaling methods including minimum-max-normalization. For effective model evaluation it was divided into more training & tested sets towards clean & organized dataset.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

iii) Feature Selection:

To find most relevant properties that strongly help among identification & classification of DDOS attacks, plant was chosen. First, fixed correlated & fruitless functions were completed using correlation analysis. features were then ranked according towards their prognosis value using a wooden-based functional techniques, such as seen in random forest & XGBoost models. In addition, most impressive collection of features was repeated using Recursive Feature Elimination (RFE). This process shrunk overfit, increased model accuracy & functional room. proposed classification model was then trained using final properties selected.

iv) Training & Testing:

The DDOS LAB dataset was divided into two subsets for training & test process: 80% for training & 20% for testing. This was done according towards standard machine learning processes towards guarantee a reliable evaluation. training model, such as random forest, XGBoost & others, was mounted for diversity. model was trained towards identify & distinguish trends between a number of DDOs attacks. performance of model was evaluated using remaining 20% data, which was set separately for testing. In order towards evaluate success of classification, F1 score, recall, accuracy & accuracy was calculated.

v) Algorithms:

Based on theorem, which facilitates freedom, naive Bayes classification is likely. It chooses square among highest probability after calculating this possibility that a sample is of each square. Due towards use & simple efficiency, this method is often used towards explore text classification & spam. In many real world applications where freedom perception cannot endure made, Bole Baye's surprisingly performs well despite view of freedom [20].

support determines best hyperplane towards divide a monitored learning system data called Support Vector Machine (SVM) in two groups. towards improve generalization of unknown data, it wants towards maximize margin between two classes. SVMs abide often used for classification tasks, such as Bidirectional science & image classification. Lesson & gene expression data abide examples of high-dimensional regions where there is Excel [19].

A statistical model for binary classification is called logistic regression. This uses a logistic function towards estimate possibility that a given entrance is of a specific class. For tasks such as risk assessment, medical diagnosis & spam detection, logistics abide easy towards understand, straight & effective. Because of efficiency & success of linear classification problems, it is widely used in many other domains [20].

towards make decisions based on functional values, decision tree produces a model that looks like a tree. A convenience -based decision is represented among each node, while potential results abide represented through branches. It is often used in classification & regression applications & is easy towards understand. Due towards its simplicity & clarity, decision trees abide used in industries such as consumer department, health care & finance [19].

A non-parametric, slow learning method for classification & regression problems is called K-Nearest Neighbors (KNN). data is classified according towards majority class in functional room of function room. Simple simplicity, effect & implementation of KNN makes it a popular choice for pattern recognition & recommended systems. It can endure ineffective among gigantic datasets, but it works well among little dataset [20].

In order towards increase accuracy, voting classifier integrates predictions from several machine learning models. average of production from majority votes or different base classifications is used towards determine which class is assigned. When a single model cannot work well, a cloth approach increases flexibility & often uses. Voting classifies is often used towards improve model generalization & reliability in classification problems [19].

4. RESULTS & DISCUSSION



"Fig.3 Output Screen"

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025



"Fig.4 Result for Malware URL Detection"

5. CONCLUSION

In this study, we proposed a double machine learning strategy towards prevent & identify IoT bottle attacks. towards detect attack, we trained a state -Ar -art deep learning model, Rachanet -18, & called it recanescan -1 model in first fold. In second fold, if scanning detection model is unable towards stop BOTT time attack, we trained a second recanet-18 model (called Resnetdos-1 model) towards identify DDOS attack. We did some experiments towards confirm efficiency of recommended recanescan-1 & recanetdos-1 model. We used scans & DDOS traffic samples from three publicly available datasets towards train ResNet-18 model (called ResNetDDoS-1 model). resulting ResNetScan-1 & ResNetDDoS-1 models were then evaluated using test set among several data sets that were not used for training. According towards experimental results, when evaluated on datasets that abide not used for training, all Res Net Scan & Res Net DDOS models—aside from suggested ResNetScan-1 & ResNetDDoS-1 model. In addition, practical results showed that suggested recanescan-1 & recanetdos-1 model performed better & better than any other model towards detect DDOS & scan attacks. Consequently, IoT is effective & reliable towards prevent & identify IoT Botnet attacks among a wide range of patterns of proposed double strategy attack.

Only thirty -three scanning types & sixty DDOS attacks abide included in current experiment. towards train proposed structure properly towards prevent & detect IoT Botnet & DDOS attacks, we intend towards cover function of more scanning & DDO's attacks in future. In addition, we can use a two -dimensional strategy proposed in an IDS towards check effectiveness of network traffic in real time.

REFERENCES

[1] I. Ali, A. I. A. Ahmed, A. Almogren, M. A. Raza, S. A. Shah, A. Khan, & A. Gani, ``Systematic literature review on IoT-based botnet attack," IEEE Access, vol. 8, pp. 212220_212232, 2020.

[2] S. Ghazanfar, F. Hussain, A. U. Rehman, U. U. Fayyaz, F. Shahzad, & G. A. Shah, ``IoT-Flock: An open-source framework for IoT traf_c generation," in Proc. Int. Conf. Emerg. Trends Smart Technol. (ICETST), Mar. 2020, pp. 1_6.

[3] M. Safaei Pour, A. Mangino, K. Friday, M. Rathbun, E. Bou-Harb, F. Iqbal, S. Samtani, J. Crichigno, & N. Ghani, ``On data-driven curation, learning, & analysis for inferring evolving Internet-of-Things (IoT) botnets in wild," Comput. Secur., vol. 91, Apr. 2020, Art. no. 101707.

[4] F. Hussain, S. G. Abbas, M. Husnain, U. U. Fayyaz, F. Shahzad, & G. A. Shah, ``IoT DoS & DDoS attack detection using ResNet," in Proc. IEEE 23rd Int. Multitopic Conf. (INMIC), Nov. 2020, pp. 1_6.

[5] S. Dange & M. Chatterjee, ``IoT botnet: largest threat towards IoT network," in Data Communication & Networks. Singapore: Springer, 2020, pp. 137_157.

[6] F. Hussain, S. G. Abbas, U. U. Fayyaz, G. A. Shah, A. Toqeer, & A. Ali, "Towards a universal features set for IoT botnet attacks detection," in Proc. IEEE 23rd Int. Multitopic Conf. (INMIC), Nov. 2020, pp. 1_6.

[7] A. O. Proko_ev, Y. S. Smirnova, & V. A. Surov, ``A method towards detect Internet of Things botnets," in Proc. IEEE Conf. Russian Young Res. Electr. Electron. Eng. (EIConRus), Jan. 2018, pp. 105_108.

[8] B. K. Dedeturk & B. Akay, ``Spam _ltering using a logistic regression model trained through an arti_cial bee colony algorithm," Appl. Soft Comput., vol. 91, Jun. 2020, Art. no. 106229.

[9] N. Vlajic & D. Zhou, ``IoT as a land of opportunity for DDoS hackers," Computer, vol. 51, no. 7, pp. 26_34, 2018.

[10] GitHub Survived Biggest DDoS Attack Ever Recorded. Accessed: May 3, 2021. [Online]. Available: https://github.blog/2018-03-01-ddosincident- report/

[11] AWS Said it Mitigated a 2.3 Tbps DDoS Attack, Largest Ever. Accessed: May 3, 2021. [Online]. Available: https://www.zdnet.com/article/awssaid-it-mitigated-a-2-3-tbps-ddos-attack-the-largest-ever/

[12] Shodan. Accessed: May 3, 2021. [Online]. Available: https://www. shodan.io/

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

[13] Censys. Accessed: May 3, 2021. [Online]. Available: https://censys.io/ [14] C. Kolias, G. Kambourakis, A. Stavrou, & J. Voas, ``DDoS in IoT: Mirai & other botnets," Computer, vol. 50, no. 7, pp. 80_84, 2017.
[15] R. Hallman, J. Bryan, G. Palavicini, J. Divita, & J. Romero-Mariona, ``IoDDoS_The internet of distributed denial of sevice attacks," in Proc. 2nd Int. Conf. Internet Things, Big Data Secur. Setúbal, Portugal: SciTePress, 2017, pp. 47_58.
[16] H.-T. Nguyen, Q.-D. Ngo, & V.-H. Le, ``A novel graph-based approach for IoT botnet detection," Int. J. Inf. Secur., vol. 19, no. 5, pp. 567_577, Oct. 2020.
[17] F. Hussain, S. G. Abbas, G. A. Shah, I. M. Pires, U. U. Fayyaz, F. Shahzad, N. M. Garcia, & E. Zdravevski, ``A framework for malicious traf_c detection in IoT healthcare environment," Sensors, vol. 21, no. 9, p. 3025, Apr. 2021.
[18] Y. Meidan, M. Bohadana, Y. Mathov, Y. Mirsky, A. Shabtai, D. Breitenbacher, & Y. Elovici, ``N-BaIoT_Network-based detection of IoT botnet attacks using deep autoencoders," IEEE Pervasive Comput., vol. 17, no. 3, pp. 12_22, Jul./Sep. 2018.
[19] H. Hindy, D. Brosset, E. Bayne, A. Seeam, C. Tachtatzis, R. Atkinson, & X. Bellekens, ``A taxonomy & survey of intrusion detection system design techniques, network threats & datasets," 2018, arXiv:1806.03517.
[20] L. Le Jeune, T. Goedemé, & N. Mentens, ``Machine learning for misuse-based network intrusion detection: Overview, uni_ed evaluation & feature choice comparison framework," IEEE Access, vol. 9, pp. 63995_64015, 2021.
[21] C. Kim, M. Jang, S. Seo, K. Park, & P. Kang, ``Intrusion detection based on sequential information preserving log embedding methods & anomaly detection algorithms," IEEE Access, vol. 9, pp. 58088_58101, 2021.
[22] V. T. Alaparthy & S. D. Morgera, ``A multi-level intrusion detection system for wireless sensor networks based on immune theory," IEEE Access, vol. 6, pp. 47364_47373, 2018.
[23] K. He, X. Zhang, S. Ren, & J. Sun, ``Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770_778.
[24] W. Wang, Y. Shang, Y. He, Y. Li, & J. Liu, ``BotMark: Automated botnet detection among hybrid analysis of _ow-based & graph-based traf_c behaviors," Inf. Sci., vol. 511, no. 2, pp. 284_296, Feb. 2020.
[25] W. Yassin, R. Abdullah, M. F. Abdollah, Z. Mas'ud, & F. A. Bakhari, ``An IoT botnet prediction model using frequency based dependency graph: Proof-of-concept," in Proc. 7th Int. Conf. Inf. Technol., IoT Smart City, Dec. 2019, pp. 344_352.