RESEARCH ARTICLE

DOI: 10.47750/jptcp.2023.30.09.030

A Ddos Attack Categorization and Prediction Method Based on Machine Learning

S.Siva saravanababu^{1*}, G.Saravanakumar²,Naveen V M³, Ajitesh kumar A S B⁴, Koushik P H⁵, Carolyne Sneha⁶, Bhuvaneswari⁷

¹Associate Professor, Department of Electronics and Communication Engineering.

²Professor, Department of Electronics and Communication Engineering (Karpagam College of Engineering)

^{3,4,5,6,7}Final Year UG Students, Department of Electronics and Communication Engineering, (Vel Tech High Tech Dr Rangarajan Dr Sakunthala Engineering College Chennai, India)

*Corresponding author: S.Siva saravanababu, Associate Professor, Department of Electronics and Communication Engineering, Email: sivasaravanababu@velhightech.com

Submitted: 22 February 2023; Accepted: 11 March 2023; Published: 30 April 2023

ABSTRACT

The most popular term for distributed network attacks is distributed denial of service (DDoS) attacks. These attacks employ certain limitations imposed on each arrangement asset, such as the design of the authorised organisation site. In this research, it is suggested that DDoS attack types be classified and foreseen using machine learning. The classification algorithms KNN and DNN are employed in this project's work. StandardScaler is used to pre-process the datasets. After using StandardScaler to remove the mean, the data are scaled to the unit variance. An evaluation of the model's performance was done using the confusion matrix created by the proposed project. For both Precision (PR) and Recall in the first classification, the KNN classifier technique is utilised (RE). The second classification makes use of the DNN classifier approach

Keywords: DDOSattack, KNNAlgorithm, DNNAlgorithm, Standard Scalar, Confusionmatrix, KNNclassifier technique, DNN classifier approach

INTRODUCTION

The cybercriminal as conduct an attack on distributed denial-of-service in the order to overload this server's infrastructure as a traffic. As a result, legitimate traffic cannot access the site, which slows it to a crawl or even crashes it. Your internet business may suffer serious harm from this kind of attack. The types of people and organisations eager to conduct this kind of strike and the DDoS causes might vary greatly. Some attacks are carried out by rate individuals or hacktivist who want to take down a company server in order to prove a point,

have fun by taking advantage of cyberweaknesses, or vent their anger. Some DDoS assaults are financially driven, such as when a rival blocks or suspends another company's online operations in order to bombard its target with User Datagram Protocol packets. By this technique, arbitrary ports on a remote machine are overloaded. A packet with the message "Destination Unreachable" is sent back in response if an application is not listening on that port as a result of this During this process, the host's resources are depleted, which may result in accessibility issues.

The icmp flood attempt like a UDP attack that sends an largest number of icmp echo request ping packets towards the packet to target resource. The overwhelming majority of packets are sent without for waiting response. This kind of attack can utilize the both inbound and outbound bandwidth, significantly shutting down the entire system, as the victim's servers rapidly attempt to react with icmp echo to ensure the data to know as weakness in the three ways of handshake which requires a SYN request to establish a tcp connection to the host to be met by an SYN Ack response to create the host, that might be launch a syn flood ddos assult.the ip packet that include the header and also cannot be a larger than 65,535 bytes. The data connectivity layer trends to be a higher layer and the largest frame size on an Ethernet network at 1500 bytes. A huge IP packet is split up into a number of smaller IP packets, or fragments, in this case. The receiving host then puts the fragments back together to make the entire packet. When purposely altering fragment content leads to an IP packet being forwarded to the destination that is larger than 65,535 bytes when put together, this is known as a "Ping of Death." As a result, the memory buffers for the packet may overflow, denying service to legitimate packets. When the slow loris attack a web server may took down another without affecting other service or ports of the target network. The slow loris does this by keeping a large number of connection open in the web server and it accomplish this by the connecting to the target computer and sends only a portion of the request. Slow loris rapidly send the new HTTP data but it never completes the request. The specified site maintains each of these false relationships and for further connection from the valid customers and when it starts to increases it will block when the maximum concurrent connection pool gradually fills. The attacker exploits Publicly Accessible Network time portal services to flood a target machine. When the query to response ratio 1:20 -1:200 or more the assault is classified as a amplification barrage so this means the any attacker with access to a list of accessible to the network time portal server may instanly launch a devastating higher bandwidth, high volume ddos attack. An HTTP deluge ddos attack targets the

webs server or application by sending request that can appear to be a normal HTTP GET ot POST request. When s compared to the other forms of attacks the HTTP requires less data since and don't involve spoofing, reflection or malformed messages to bring down the targeted server, application r website. The assault is most effective when the server or program is compelled to provide every request the maximum amount of resources that are available and the board meaning of the machine learning branch of the AI is a system capacity to give the human behavior. AI system are used to tackle the complicated tasks in a manner similar to how the people handle the challenges and also it have sufficient amount of organized varied data is required for a good ML answer including ML. All businesses now have the accesses to vast amount of data of their customer in todays online first society.

LITERATURE SURVEY

Kshira Sagar Sahoo et al [2020] proposed a The Software Defined Network (SDN) architecture is gained popularity among network operators as it provides them with increased control over their network's infrastructures. In this paper, we propose a method for detecting attack traffic using the centralized control feature of SDN. To reduce the dimension of feature vectors, we employ Kernel Principle Component Analysis in our Support Vector Machine model. We also optimize various SVM parameters using a Genetic Algorithm (GA). To address the issue of noise caused by feature differences, we propose a new kernel function called N-RBF. Our results shows the proposed model that outperforms single SVM on the terms of classification of the accuracy and generality and further more the model may be incorporated into the controller to provide the security rules that can protect against probable attackers threats. The possibility of creating more complex algorithm by fusing KPCA technique is not investigated in this article.

Ankit Agarwal et al [2021] developed Deep learning algorithms are effective in categorizing both normal and large attacked data. We present a unique feature selection to optimized approach

to handle the issue of feature selection. A deep neural network approach for effectively on ddos assaults is presented in the paper and we use minimum to maximum to normalization to bring the all input values inside in the given range. The suggested FSWOA and then used to choose the optimal collection of the characteristics for the class cation procedure and given the severity of DDoS in large multinational attacks corporations, developing research on attack prevention models is beneficial. However, this paper does not consider IDS schemes for detecting novel attacks as individual instances.

Jin Ye et al [2018] The emergence of Software-Defined Networks (SDN) has led to the development of novel methods in this area. One such solution is to use deep learning systems to simulate attack behavior based on data received from SDN controllers. For this study, we will use the Mininet and Floodlight simulation platforms to create an SDN environment. We extract the 6tuple metric from the switch flow table and use the SVM classification method to develop a DDoS attack model. To detect DDoS attacks, we use the SOM method to extract DDoS attack statistics. This method has a high detection rate with low resource consumption. However, it suffers from some hysteresis in detection, and the attack behavior may not be identified in a timely and accurate manner

Jesús Arturo Pérez-Díaz et al [2020] The adoption of IPv6 solved the problem of address exhaustion in IPv4, but it also introduced the usage of ICMPv6 messages in newer technologies like the Neighbor Discovery Protocol. This article examines the IDSs Used to identify and classify ICMPv6 based Dos and DDoS threats into single and hybrid classifiers. Blockchain technology is integrated into the Collaborative IDS architecture, using ensemble framework to solve the problem of false positive alarms in ICMPv6 based DoS and DDoS detection. This is the first review study in this area that focuses exclusively on IDS based on ML techniques with the aim of improving researchers' knowledge of developing IDS models based on ensemble learning. The paper also discusses the relevance of blockchain as a research topic in this field, and it suggests that the number of false-positive alerts from DoS and

DDoS attack detection based on ICMPv6 should typically decrease.

Neha Agrawal et al [2019] It has been observed that DDoS attacks in cloud environments involve flooding victim servers with high-volume malicious traffic, which makes them easily detectable. This study provides a complete taxonomy of all potential cloud DDoS attack variants and insights on their characterization, prevention, detection and mitigation strategies. Its purpose is to encourage cloud security researchers to develop viable DDoS mitigation systems. In addition, the report also discusses research gaps and issues as well as future research topics. Preventing and containing such attacks is difficult due to the sheer volume of attack traffic involved.

[2018] Kiwon Hong el al A slow HTTP distributed denial of attack has been suggested as a protection approach. This makes the web server unusable and difficult to detect on the network because it mimics real client traffic. To identify and neutralize such threats, defense mechanisms are network-based and use Software-Defined Networks (SDN). Simulation results show that the proposed strategy was successful. In the proposed technique, SHDA accepts and processes suspicious incomplete HTTP requests on behalf of web servers. This content does not contain identified plagiarism.

Shi Dong et al [2020] Distributed Denial of Service assaults still lack a reliable defense mechanism after many years of significant network availability problems. Software Defined Networking (SDN), however, presents a novel perspective on DDoS assault defense. This study recommends two strategies for identifying SDN DDoS attacks. One method employs the DDoS attack's intensity to identify it, while the other method makes use of an improved K-Nearest Neighbors (KNN) machine learning approach. The suggested techniques outperform previous methods in detecting DDoS attacks, according to theoretical analysis and experimental results on datasets. Their efficiency needs to be improved in order for DDADA and DDAML to be useful in a real-world SDN situation.

Mitali Sinha et al [2021] Network-on-chip architectures are vulnerable to flood-based

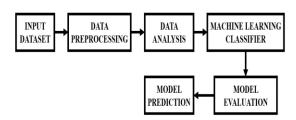
denial-of-service attacks due to their shared nature and open access to all on-chip modules. Such attacks are launched via malicious intellectual property embedded in a system-onchip flooding his NoC with worthless data and drastically reducing bandwidth. There is a possibility in this document, Sniffer presents her MIP localization method based on machine learning. It accurately tracks attack paths with minimal overhead and uses collaborative decision making to identify her MIPs. Experimental results show that Sniffer can find his MIP with high accuracy and less effort however in the context of NoC based systems, attack source localization is an area that requires further investigation.

Mohammad Tayyab et al [2020] This article discusses how the adoption of IPv6 addressed an issue of IPv4's address exhaustion, which resulted in the use of ICMPv6 messages in newer technologies like the Neighbor Discovery Protocol. The paper extensively analyzes the intrusion detection systems used to detect ICMPv6 based Dos and Distributed Denial of Service attacks dividing them into both hybrid and single classifiers. It also proposes using blockchains in the Collaborative IDS architecture based on the ensemble framework, to tackle one of the challenges in detecting ICMPv6 based Dos and DDos threats. The paper is the first review study that mainly focuses on IDSs using machine learning approaches in this area, with the aim of encouraging researchers to develop group learning based IDS models. Additionally, the article brings up the topic of false-positive alerts from DoS and Distributed Denial of Service attack detection based on ICMPv6, which could be reduced with the use of blockchains.

PROPOSED METHOD

In recent years, DDoS attacks have received increased media coverage, prompting researchers to investigate potential solutions to network security issues. One approach that has garnered significant attention is Software Defined Networking (SDN). While SDN presents many benefits, its architecture is fundamentally different from traditional networks, making it vulnerable to DDoS attacks, particularly the SDN

controller.To exploit vulnerabilities in the network header fields, researchers developed an SDN network scanning tool capable identifying SDN networks. By querying the controller through the data path and flooding it with flow setup requests, attackers can overload the controller and cause it to break down. As a a DDoS detection technique necessary. However, the deployment of multiple controllers can exacerbate the problem by resulting in cascading controller faults. This highlights the need for an effective DDoS detection mechanism that can mitigate the effects of such attacks.Researchers have proposed several methods for detecting DDoS attacks in SDN networks. For instance, machine learning techniques can be used to detect anomalous traffic patterns that could indicate the presence of a DDoS attack. Ensemble learning-based IDS models have also been developed and integrated with blockchain technology to enhance detection capabilities.In conclusion, the unique of architecture SDN networks presents significant challenges in protecting against DDoS attacks. Effective detection mechanisms are necessary to mitigate the impact of such attacks, and researchers have proposed several techniques to achieve this goal. However, continued research and development in this area are necessary to improve the security of SDN networks and protect against evolving threats.



The numerous types of DDoS attacks were identified and predicted using a machine learning method in this study. An analysis of the incoming dataset takes place during the first stage of data pre-processing. Data pre-processing procedures are used to deal with the unnecessary data in the dataset. Researching the observed data is part of the data analysis stage after data pre-processing. The following machine learning classifier method is acceptable for the data. The KNN and DNN algorithms are used in the machine learning

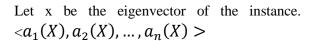
classifier approach. The first classification employs the KNN classifier technique for both Precision (PR) and Recall (RE). Precision (PR) and Recall are both classified using the DNN classifier technique in the second classification (RE). The data are then assessed to produce a forecast result. The dataset used must be considered if one wants to guarantee the accuracy of intrusion detection systems. Secure network resilience is required due to the exponential growth of networks and applications in modern times. By choosing the appropriate learning and testing datasets, it might be done. Data preprocessing is a crucial step in data preparation that involves any processing done on raw data to prepare it for another data processing operation. Because excellent data is more important than good models, it is regarded as an essential initial stage in the data mining process because data quality is so crucial, businesses and people spend a significant amount of effort cleaning and preparing data for modelling real-world data has various quality concerns, including: B. Loudness, unreliability, and incompleteness they might also be lacking certain or relevant qualities, as well as missing, inaccurate, or erroneous values. As a result, it is critical to increase the quality of data preparation in order to assure data consistency this may be accomplished by eliminating duplicates and anomalies, standardising data for comparison, and increasing results quality. Data cleaning is the process of removing or correcting incorrect, corrupt, erroneous, duplicate or incomplete data from a dataset that bringing together numerous data sources improves the chances of duplicating or mislabeling data.

Data preparation is an essential step in the data mining process, as not all real-world data is exhaustive, accurate, correct, consistent, and pertinent. The first and most important step in data preparation is cleaning up the data. This process involves several steps, including maintaining consistency in the data across all values, checking for null or missing values, and handling outliers to reduce noise in the data. Evaluation techniques are also employed to identify the best-fit algorithm that will produce the desired results for the given dataset. Accuracy is the primary consideration when using machine

learning to address various issues. An ML problem can be solved in various stages, including data collection, problem definition, brainstorming with the available transformation, preprocessing, training the model, and evaluation. The evaluation stage is the most critical since it allows us to gauge how accurately the model predicts the future. The effectiveness and use of the ML model are decided based on accuracy metrics. Many data models have been developed to meet various data analysis needs, as data analysis varies from company to company. needs. The division of data analytics known as predictive modelling employs probability and data mining to forecast outcomes. The number of predictors that are very helpful in predicting future decisions is used to build each model. An analytical model is created once the data for a given predictor is received. Assumptions are an essential component of any data model; they not only make the model simple but also improve predictions. As a result, Parametric and non-parametric machine learning algorithms differ in their approach to modeling data. Parametric algorithms make assumptions about the functional form of the data, which allows them to simplify the problem and estimate model parameters based on a fixed set of variables. On the other hand, non-parametric algorithms do not make strong assumptions about the underlying data distribution and instead rely on flexible models that can adapt to changing patterns in the data. Parametric models have a fixed number of parameters and assume that the data follows a specific distribution, which limits their flexibility in handling complex data. Nonparametric models, on the other hand, do not have a fixed number of parameters and can adapt to new data patterns, making them suitable for handling more complex and diverse data.In summary, parametric models assume a specific form for the data and rely on fixed parameters to estimate the model, while non-parametric models are more flexible and can adapt to changing patterns in the data, ML algorithms can pick up any functional form of training data. Without any prior knowledge, non-parametric models fit the vast amount of data well.

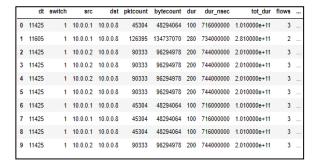
Equations

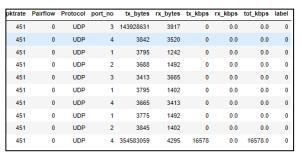
A nearest neighbor approach is used for prediction because objects with similar predicted values are close together this approach finds the k closest points to the unknown sample in the multidimensional space Rn and determines the class of the unknown sample based on the categories of these k points. This is called the k nearest neighbors this approach assumes that the instance is a point in dimensional space and that the nearest neighbor of the instance is found using standard Euclidean distance.

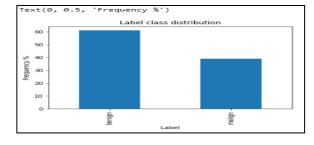


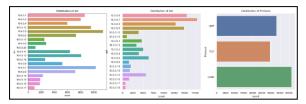
The formula to calculate the distance between two instances is provided as follows: defined $asd(X_i, Y_i)$,

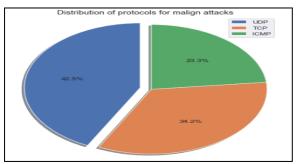
$$d(X_i, Y_j) = \sqrt{\sum_{r=1}^{n} (ar(X_i) - ar(x_j))2}$$

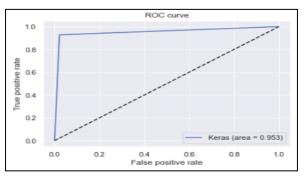






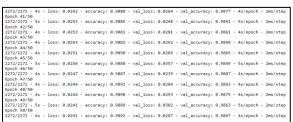


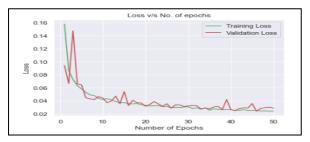


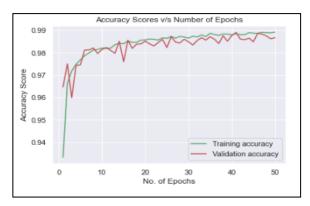


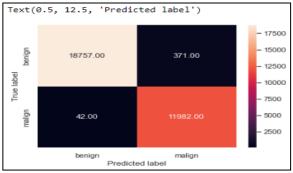
KN











RESULT AND DISCUSSION

A methodical technique to DDOS attack detection is presented in our paper. Initially, we used the Australian Centre for Cyber Security (ACCSUNSW-nb15)'s dataset from the GitHub repository, which contains statistics on DDOS assaults. We did data wrangling using Python and Jupyter notebook, splitting the dataset into dependent and independent classes, which were afterwards normalised for algorithmic analysis. Using deep the Random Forest and XGBoost algorithms, our suggested supervised machine learning approach produced predictions and classification results. For the first classifier, our model has an average accuracy of 89% and for the second, it has an average accuracy of 90%. Our model greatly increased flaw determination accuracy as compared to previous research. Going ahead, our goal is to offer a quicker, easier-to-use alternative to learning computations that nevertheless delivers superior outcomes.

For both labelled and unlabeled datasets, we will also investigate unsupervised learning strategies to increase detection accuracy.

We also intend to look at how well nonsupervised learning systems can identify DDOS assaults.

CONCLUSION

In recent years, distributed denial of service has assaults have been among the most common and serious cyberattacks. DDoS attacks aim to stop a victim's network from operating normally by flooding it with traffic and making it unreachable to users. Attackers can initiate DDoS assaults on specific systems using botnets or other resources, frequently with the intention of harming the victim's finances or reputation. This research suggests a thorough simulation of a machine learning approach to identify and forecast various forms of DDoS assaults in order to enhance detection and understanding of DDoS attacks. Pre-processing, data analysis and categorization, performance measurements simulation's three core parts. The incoming traffic data is prepared for analysis by the preprocessing module. This entails putting all incoming traffic parameters on a common scale to make it easier to use various machine learning methods. Also, the data is cleaned and filtered to eliminate any noise or unrelated information that can obstruct the study. The real machine learning algorithms are used on the pre-processed data in the data analysis and classification module. The data is divided into several DDoS attack groups using logistic regression and KNN, two classification techniques. While KNN is a nonparametric MLalgorithm used classification and regression, logistic regression is a statistical technique used to anticipate a binary result. The efficacy of the machine learning model is assessed via the performance metrics module. This requires measuring the model's accuracy, precision, recall, and F1-score. By adjusting the settings of the preprocessing and machine learning algorithms, the efficacy of the model is meant to be optimized and this was the main goal of this research is to increase our capacity to categorise various DDoS assault types using machine learning techniques. We may be able to better understand the patterns and traits of DDoS attacks by replicating these methods, which will enable us to create preventative and mitigation tactics that are more effective.

REFERENCES

- N. Agrawal and S. Tapaswi, "Defense Mechanisms Against DDoS Attacks in a Cloud Computing Environment: State-of-the-Art and Research Challenges," in IEEE Communications Surveys & Tutorials, vol. 21, no. 4, pp. 3769-3795, Fourthquarter 2019.
- K. Hong, Y. Kim, H. Choi and J. Park, "SDN-Assisted Slow HTTP DDoS Attack Defense Method," in IEEE Communications Letters, vol. 22, no. 4, pp. 688-691, April 2018.
- S. Dong and M. Sarem, "DDoS Attack Detection Method Based on Improved KNN With the Degree of DDoS Attack in Software-Defined Networks," in IEEE Access, vol. 8, pp. 5039-5048, 2020.
- Y. Xu, H. Sun, F. Xiang and Z. Sun, "Efficient DDoS Detection Based on K-FKNN in Software Defined Networks," in IEEE Access, vol. 7, pp. 160536-160545, 2019.
- S. Ali and Y. Li, "Learning Multilevel Auto-Encoders for DDoS Attack Detection in Smart Grid Network," in IEEE Access, vol. 7, pp. 108647-108659, 2019.
- M. Sinha, S. Gupta, S. S. Rout and S. Deb, "Sniffer: A Machine Learning Approach for DoS Attack Localization in NoC-Based SoCs," in IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 11, no. 2, pp. 278-291, June 2021.
- M. Tayyab, B. Belaton and M. Anbar, "ICMPv6-Based DoS and DDoS Attacks Detection Using Machine Learning Techniques, Open Challenges, and Block chain Applicability: A Review," in IEEE Access, vol. 8, pp. 170529-170547, 2020.
- H. A. Alamri and V. Thayananthan, "Bandwidth Control Mechanism and Extreme Gradient Boosting Algorithm for Protecting Software-Defined Networks Against DDoS Attacks," in IEEE Access, vol. 8, pp. 194269-194288, 2020.
- 9. Aljuhani, "Machine Learning Approaches for Combating Distributed Denial of Service Attacks in Modern Networking Environments," in IEEE Access, vol. 9, pp. 42236-42264, 2021.
- R. Biswas, S. Kim and J. Wu, "Sampling Rate Distribution for Flow Monitoring and DDoS Detection in Datacenter," in IEEE Transactions on Information Forensics and Security, vol. 16, pp. 2524-2534, 2021.

- K. S. Sahoo et al., "An Evolutionary SVM Model for DDOS Attack Detection in Software Defined Networks," in IEEE Access, vol. 8, pp. 132502-132513, 2020.
- Agarwal. A, Khari. M., & Singh, "R. Detection of DDOS attack using deep learning model in cloud storage application," Wireless Personal Communications, pp. 1-21, 2021.
- W. Zhijun, X. Qing, W. Jingjie, Y. Meng and L. Liang, "Low-Rate DDoS Attack Detection Based on Factorization Machine in Software Defined Network," in IEEE Access, vol. 8, pp. 17404-17418, 2020.
- B. Nugraha and R. N. Murthy, "Deep Learning-based Slow DDoS Attack Detection in SDN-based Networks," 2020 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN), Leganes, Spain, 2020
- J. A. Pérez-Díaz, I. A. Valdovinos, K. -K. R. Choo and D. Zhu, "A Flexible SDN-Based Architecture for Identifying and Mitigating Low-Rate DDoS Attacks Using Machine Learning," in IEEE Access, vol. 8, pp. 155859-155872, 2020.
- 16. N. Agrawal and S. Tapaswi, "Defense Mechanisms Against DDoS Attacks in a Cloud Computing Environment: State-of-the-Art and Research Challenges," in IEEE Communications Surveys & Tutorials, vol. 21, no. 4, pp. 3769-3795, Fourthquarter 2019.
- 17. K. Hong, Y. Kim, H. Choi and J. Park, "SDN-Assisted Slow HTTP DDoS Attack Defense Method," in IEEE Communications Letters, vol. 22, no. 4, pp. 688-691, April 2018.
- S. Dong and M. Sarem, "DDoS Attack Detection Method Based on Improved KNN With the Degree of DDoS Attack in Software-Defined Networks," in IEEE Access, vol. 8, pp. 5039-5048, 2020.
- 19. Y. Xu, H. Sun, F. Xiang and Z. Sun, "Efficient DDoS Detection Based on K-FKNN in Software Defined Networks," in IEEE Access, vol. 7, pp. 160536-160545, 2019.
- S. Ali and Y. Li, "Learning Multilevel Auto-Encoders for DDoS Attack Detection in Smart Grid Network," in IEEE Access, vol. 7, pp. 108647-108659, 2019.