'National Conference on Advancements in Computer Science and Engineering'
International Journal of Scientific Research in Computer Science,
Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

A Unified Intelligent and Time-Efficient DDoS Attack Prediction using Tree Based Algorithm

Dr. Shaamili Varsa G V[1], S. Padmanabhan[2], Vidya Sagar T[2], Soloman Richard[2]
[1]Assistant Professor, [2]Student
Department of Computer Science & Engineering, SRM Institute of Science & Technology, Ramapuram, Chennai, Tamil Nadu, India

ABSTRACT
DDoS attack has been a significant problem for decades, causing severe disruptions to network availability, yet there is no defense against it. However, with the availability of SDN there is a chance to defend against DDoS attacks in SDN. We propose two methods for identifying DDoS attacks [5]. One method detects the DDoS attack by its degree and the other uses a machine learning based enhanced tree-based algorithm for discovering the attack [5]. The analysis of datasets and experimental results shows our work is efficient than other existing systems.

Keywords- Machine Learning, Dos/DDOS attacks, SDN (Software Defined Network), tree-based algorithm

I. INTRODUCTION

Over the last few years, DDoS assaults has received a lot of attention online. The SDN concept has been introduced in the recent years and has been researched widely [2]. SDN has 3 main parts called Application Control and Data Plane. DDoS attacks can still abuse the existence of SDN due to the difference in characteristics/architecture between SDN and traditional network The component most vulnerable to DDoS attacks is the SDN controller in particular. Many authors have researched about secondary/multiple controllers' concept for minimizing the DDoS attacks in SDN, but still the use of multiple controllers couldn't solve the problem and leads to failure of both the controllers due to the lack of DDoS detection mechanism. SDN is newer approach to manage the LAN and WAN infrastructure of any modern enterprise network and it has been popular lately among the researchers as it provides a programable and dynamic network switching mechanism unlike, Traditional network follows an 2/3 tier architecture and mostly performs OSPF protocol for transmission. In SDN, Control and data plane is separated, so that it may run on less costly hardware, SDN switches simply forward packets to the subsequent devices on the Data plane. The control plane manages the network's overall behavior, including routing protocols, network policies, and network security. It is essentially the "brains" of the network, and it makes decisions about how to handle network traffic based on high-level policies and goals. Meanwhile, the data plane is in charge of handling actual network traffic and handling the low-level specifics of network connection, such as packet forwarding, quality of service (QoS) handling, and

network address translation (NAT). The seller often provides the software. Therefore, network managers must become knowledgeable about various technologies from various suppliers and setup each device for routing.

II. RELATED WORK

In 2019, Shan Ali and Yuancheng Li put forward a proposal that employs machine learning for efficient observation of DDoS attacks in smart grid networks. The technique uses multi-level autoencoder-based feature learning, which combines both shallow and deep unsupervised auto encoders to generate potent features. These features are then aggregated using the Multiple Kernel Learning (MKL) algorithm to create a final unified recognition model. To evaluate the proposed method, the authors conducted experiments using two DDoS attack databases and their subsets, comparing the results with six other techniques. The proposed approach demonstrated superior prediction accuracy and proved to be more effective in detecting DDoS attacks in smart grid networks than comparable methods. The authors suggest that their work can be implemented in a runtime environment to provide protection against future DDoS attacks [6].

In 2021, Abimbola O. Sangodoyi and co-authors conducted a study on flood attack detection and classification in SDN. They used Low Orbit Ion Cannons (LOIC) to mimic DDoS flood assaults while simulating the SDN paradigm on Mininet. The findings indicated that machine learning techniques were useful for identifying and categorizing DDoS flood assaults, with CART showing the greatest performance in terms of prediction speed, accuracy, and dependability. The authors concluded that machine learning- based DDoS flood attack detection and classification could be implemented using popular supervised learning methods, and that CART was the most suitable method due to its stability and efficiency [7].

In 2020, Bruno Martins Rahal, Aldri Santos, and Michele Nogueira investigated the challenge of detecting and predicting DDoS attacks in a distributed network environment effectively and efficiently. They proposed a two-tier distributed architecture that integrates and analyzes both local and Internet traffic using predictive techniques to address this issue. The architecture's objective is to promote decentralized and integrated network traffic analysis while leveraging predictive methods like identifying early signs of large-scale DDoS attacks to initiate bot detection techniques [8].

To identify and anticipate distributed denial of service (DDoS) assaults, Seongyun Seo, Sungmin Han, Janghyeon Park, Shinwoo Shim, Han-Eul Ryu, Byoungmo Cho, and Sangkyun Lee presented a distributed architecture in 2021. Predictive technology is used in this architecture to spot early warning signs of future DDoS assaults and to find the individual bots that make up the botnet. The design uses unsupervised statistical learning to identify impending DDoS assaults and is based on the metastability theory. Grouping network devices according to the cause-and- effect connection between them and the characteristics derived from traffic is used for carrying out botnet identification The findings show that the suggested architecture can successfully and accurately identify the robot dataset with 99.9% accuracy when tested using various datasets. Decentralized and integrated analysis of local data is possible with the suggested architecture [9].

To effectively identify and categorise Distributed Denial of Service (DDoS) assaults, Yuanyuan Wei, Julian Jang- Jaccard, Fariza Sabrina, Amardeep Singh, Wen Xu, and Seyit Camtepe developed a mixture approach dubbed AE-MLP in 2021. Due to the abundance of network data, including malicious DDoS payloads, conventional low- level machine, learning-based approaches are inefficient at identifying DDoS attacks. The AE-MLP model combines the Autoencoder (AE) and Multi-layer Perceptron Network (MLP) deep learning

models to extract pertinent information and categorise attacks into several DDoS attack categories. The AE component of the model offers effective feature extraction that automatically identifies the most crucial feature sets, while the MLP component uses compressed and reduced feature sets produced by AE as inputs to categorise assaults. The outcomes of experiments show a high [10].

In their research article published in 2021, Marinos Dimolianis, Adam Pavlidis, and Vasilis Maglaris suggested a creative defence strategy against DDoS assaults. The suggested method makes use of signature- based filtering rules as opposed to source IP-based filtering rules, which may have scale problems. To speed up packet processing, the approach makes use of XDP middleboxes, which serve as programmable deep packet inspectors. The method pulls certain combinations of key packet characteristics from network data and feeds them to supervised machine learning algorithms in order to detect dangerous activity. The communication is subsequently categorized as harmful. For the purpose of generating a small collection of filtering rules and accelerating mitigation, malicious signatures go through a customized reduction procedure. The suggested scheme's effectiveness was assessed for signature classification precision [11].

A Software-Defined Networking (SDN) DDoS mitigation strategy with a bandwidth management mechanism and an Extreme Gradient Boosting (XGBoost) algorithm was put forth by Hassan A. Alamri and Vijey Thayananthan in 2020. The XG Boost algorithm is activated by the bandwidth control mechanism on threshold breaches using an adaptive thresholding approach based on several bandwidth profiles. Network traffic is classified as normal or abnormal using the XG Boost algorithm. The suggested technique has a positive percentage of 0.0002 in SDN and a 99.9 accuracy rate for detecting DDoS assaults. The technology lowers the packet loss ratio and effectively utilizes network resources[12].

In 2019, Luis A. Trejo and a group of researchers developed a platform called DNS-ADVP, Is designed to find and alleviate ongoing DNS DDoS attacks. This platform contains a one-class categorization that may notify users of unusual behavior and employs a visual model to evaluate the present condition of traffic in an authoritative DNS server. DNS-ADVP uses a classification method that continuously updates normal operation. The platform has been successfully tested in synthetic attacks, achieving an area under the curve of 83%, and currently used for monitoring of a real authoritative DNS server. DNS-ADVP is more efficient than traditional methods, and reduces the workload on human operators while improving traceability. The authors conducted a study of existing techniques that mitigate DNS DDoS attacks, including a UDP rule limiting the number of requests to the same IP address, Response Rate Limiting (RRL), and the technique proposed by Kambourakis. et al. The authors' new visual model enables quick interpretation of DNS traffic and detects potential deviations using visual traffic lights[13].

III. EXISTING SYSTEM

The author used an outdated KDD dataset in the current system and proposed a structured framework to predict DDoS attacks [5]. To detect the latest/current state of DDoS attack, we should work on the latest dataset. The existing system uses various techniques for DDoS attack prediction. The GitHub repository's UNWS-np dataset was downloaded, and Python was utilized as a virtual environment. In addition, the precision and reliability of the findings can be guaranteed by using the most recent datasets and creating a complete structure for DDoS attack prediction [14]. Organizations may receive timely information on prospective attacks as a result, enabling them to take preventative action to avoid network outages and any data breaches. The

suggested method effectively detects and categorizes DDoS assaults using machine learning algorithms, with promising results. This research can help to improve the security of cloud environments and vital networks because DDoS assaults continue to pose a serious danger to them. The accuracy of defect determination was significantly improved, with approximately 80% and 75%, respectively, by comparing our work to existing research. By effectively identifying and categorizing DDoS attacks, the suggested method in the current study can assist organizations in enhancing their network security. The method can manage large and complex data by utilizing machine learning methods like Random Forests and XG Boost, making it successful in identifying different forms of DDoS attacks [16].

IV. PROPOSED SYSTEM

The proposed system proposes two approaches to finding DDoS attacks in SDN.F First approach involves determining the level of DDoS attack based on its intensity. while the second approach employs a tree- based algorithm based on machine learning (ML). Initially, data analysis occurs as it involves two main steps, data collection and data integration. In data collection phase, data is obtained from real-world sources, which may be incomplete, noisy or redundant. To smooth the data, data cleaning techniques such as removing redundancy, filling missing values and smoothing data are used. Data taken from multiple sources (eg files, databases, etc.) are integrated into a single source. After data collection and integration, data reduction and transformation are performed to make the data useful for further processing. Data preprocessing removes inconsistent and noisy data to ensure accurate data analysis. Feature subsets of datasets are primarily determined by filters that depend on the size of the dataset. Many machine learning algorithms like Random Forest (RF), Relief-F etc. are used to evaluate the data type. Choosing the right subset is based on consistent criteria and can be a difficult task. Techniques such as cross validation filter (CRV), ensemble filter (EF) and partition filter (PF) are used to solve this. CRV divides the functions into subsets and each subset is tested for performance. The bad features are filtered out and the best feature is selected. On the other hand, PF divides the entire dataset into parts and selects the part where the model works best.

4.1. Algorithms Used

1. Random Forest Classification Algorithm
2. Isolation Forest
3. The Relief-F algorithm

4.2. Advantages of Proposed system

i. The Random Forest Classification Algorithm is advantageous because it can estimate missing data effectively, handle large datasets with thousands of input variables, and run efficiently

ii. The Isolation Forest Algorithm is advantageous because it works well with irrelevant features, has reduced computational times, and is easily scalable to large and high-dimensional datasets.

iii. The Relief-F algorithm has several advantages like it is computationally efficient, effective for high-dimensional data, Not influenced by redundant features, can handle continuous and discrete data as a whole.

4.3. System Architecture

The system architecture proposed in the paper "A Unified Intelligent and Time-Efficient DDoS Attack Prediction using Tree Based Algorithm" consists of several components Here is an overview of the architecture:

4.3.1. Data pre-processing:

The processes involved in cleaning data before it is utilized to build forecasting models. The data used is connected to the production of wind energy, and it has been noted that the data includes missing entries, insufficient data, anomalies, and non-normal operating situations. These problems are addressed by cleaning the data according to predefined procedures. First, the 10- minute sample cycle ignores records with missing entries (NAs). Second, records with partial entries (IN), which include one or more missing signal values, are also removed.

Moreover, observations with states that do not correspond to normal operational conditions (NNO) are also excluded. This involves examining the value of the state variables and identifying abnormal conditions, such as free rotation of wind turbines without connection to the grid or derated operation. Finally, outliers are dealt with by combining wind power observations that match to the exact same wind speed and creating a box plot. Then, outliers that have generated electricity outside the box plot's whiskers are eliminated. In more detail, observations outside the range (Q 3IQR, Q + 3IQR), where Q is the average of the information and IQR is an interquartile range, are eliminated.

4.3.2. Model Learning

Ensemble learning is a machine learning approach that aims to enhance predictive excecution by merging predictions from multiple models. Bagging, stacking, and boosting are the three main types of ensemble techniques. Data sampling can be done using either bagging or pasting. Boosting, on the other hand, trains models sequentially, with each new model building on the previous one. Various metrics can be used to evaluate the performance of multi-label classifiers, including accuracy, Hamming loss, precision, recall, and F1-Score. Accuracy measures the number of correct predictions out of all classified instances, while Hamming loss considers both prediction errors and missing errors.

4.3.3. Load dataset

The load dataset would typically consist of features that capture the behaviour of network traffic, such as packet size, packet rate, and protocol type. The models would then trigger an action to prevent the attack from reaching its target, such as blocking traffic from the source IP address or dropping packets that meet certain criteria. The dataset can also be taken from real world sources, which is raw and unprocessed. The dataset consists of DDoS Attack information such as Name, class type, Count, protocol type, duration, Service flag etc.

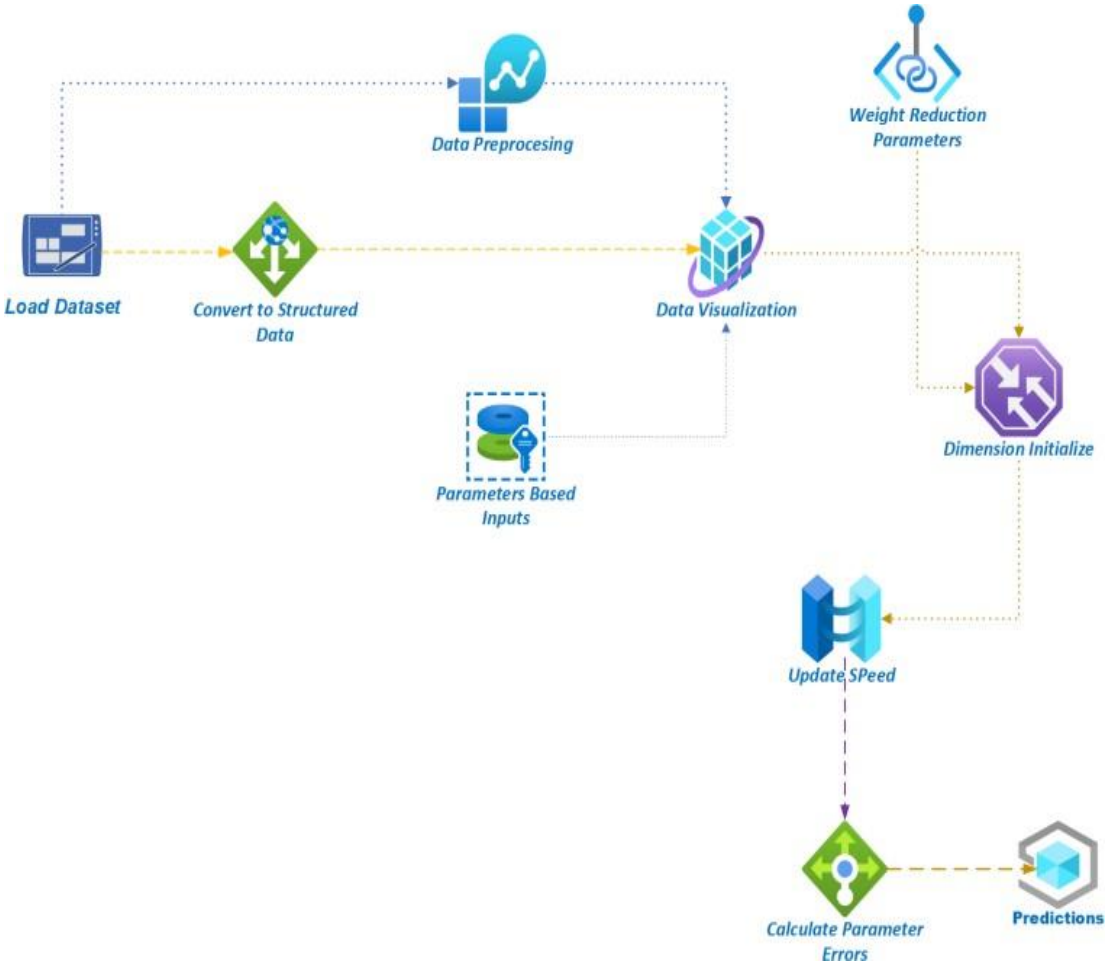

Figure 1. architecture diagram

4.3.4. Data visualization

Data visualization can help identify patterns and anomalies in traffic data that may be difficult to detect through other means. By using visualizations in combination with machine learning algorithms like Random Forest Classification and Isolation Forest, it is possible to build effective DDoS prevention systems. Data visualization is a process where the input data is analyzed visually and illustrated using charts or graphs. The techniques of data visualization have simplified the plotting of graphs for large datasets.

4.3.5. Random Forest Classification module

The objective of this module is to apply random forest classification algorithm to differentiate between normal and malicious network traffic. This process will involve hyperparameter tuning and performance evaluation of the classifier.

4.3.6. Isolation Forest module

This module would implement the Isolation Forest algorithm to detect anomalies in network traffic that may be indicative of a DDoS attack. It may involve setting hyperparameters and evaluating the performance of the anomaly detector.

V. RESULTS & DISCUSSIONS

The DDoS attack prevention of Random Forest Classification Algorithm and Isolation Forest Algorithm can involve analysis of the proposed approach. This includes comparing the precision, accuracy, recall and F1 scores of the Random Forest and Isolated Forest algorithms with other existing approaches. The discussion includes an assessment of the limitations and strengths of the proposed approach. In addition, the topic could investigate the impact of the process of weighting parameters on the performance of Random Forest and Isolated Forest algorithms to prevent DDoS attacks. Here, we analyze the effect of different weight parameters on the accuracy and efficiency of the algorithms. In the discussion, the trade-off between accuracy and computational complexity could also be considered when choosing the optimal weight.

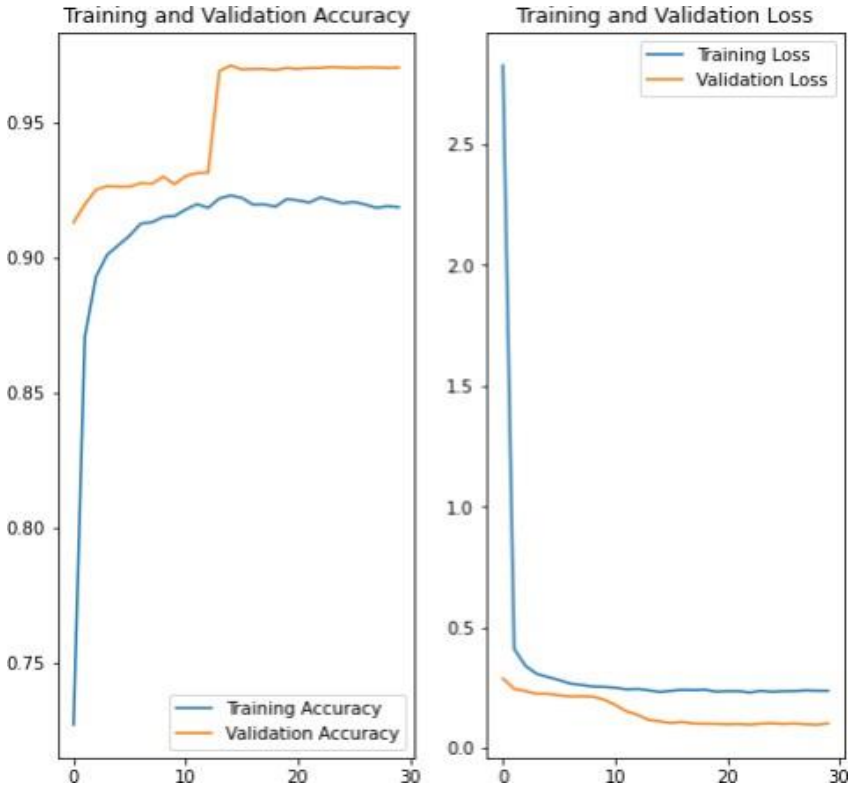

Figure 2. Difference in Training and Validation accuracy and loss

```
Confusion matrix:
 [[1478    5    0    0   18]
 [   4 3313    0    0   29]
 [   5    0    0    0    0]
 [ 109    1    0    0    0]
 [   1    0    0    0  410]]
              precision    recall  f1-score   support

           0       0.93      0.98      0.95      1501
           1       1.00      0.99      0.99      3346
           2       0.00      0.00      0.00         5
           3       0.00      0.00      0.00       110
           4       0.90      1.00      0.94       411

    accuracy                           0.97      5373
   macro avg       0.56      0.59      0.58      5373
weighted avg       0.95      0.97      0.96      5373
```

Figure 3. confusion matrix of the test data for identification of the model performance.

VI. CONCLUSION

The DDoS assault poses a serious risk to the SDN network's network security, and preventing it requires being able to detect it. However, current DDoS attack detection techniques have poor accuracy and are susceptible to other variables. We have succeeded in the following to overcome these problems:
First, when a DDoS assault targets the SDN controller, four features—flow length, duration, size, and ratio—have been examined. Degree of assault is a brand-new idea that we've proposed and it may be used to identify DDoS attacks. In comparison to existing solutions, our suggested algorithms have demonstrated greater detection rates and improved identification of DDoS assaults.

VII. REFERENCES

[1]. H. Zaki, N. Zaki, and A. Elazab, "Machine learning- based classification and prediction technique for DDoS attacks," IEEE Access, vol. 9, pp. 32410-32421, 2021.

[2]. Shin, S., & Gu, G. (2013, August). CloudWatcher: Network security monitoring using OpenFlow in dynamic cloud networks. In 2013 IEEE Conference on Communications and Network Security (pp. 78-86). IEEE.

[3]. M. F. Abdollah, N. M. Tahir, and R. Salleh, "Software-Defined Networking (SDN): A Survey," in 2016 International Conference on Recent Trends in Information Technology (ICRTIT), Chennai, India, 2016, pp. 1-6, doi: 10.1109/ICRTIT.2016.7569502.

[4]. N. M. Khan, A. F. Ahmed, S. A. Madani, and N. A. Alrajeh, "Software-defined networking: A comprehensive survey," Journal of Network and Computer Applications, vol. 75, pp. 252-276, Nov. 2016.

[5]. Ismail, M., Hassan, S. A., Abdullah, A. H., & Saad, N. M. (2021). Machine Learning-Based Classification and Prediction Technique for DDoS Attacks. IEEE Access, 9, 52140-52150.

[6]. Ali, S., & Li, Y. (2019). A multi-level autoencoder- based feature learning approach for efficient detection of DDoS attacks in smart grid networks. IEEE Access, 7, 83657-83666. doi: 10.1109/ACCESS.2019.2926946

[7]. Sangodoyi, A. O., Oluwafemi, A., Adewale, A. O., & Oluyomi, A. O. (2021). Flood attack detection and classification in software-defined networking: A machine learning approach. Journal of Ambient Intelligence and Humanized Computing, 12(7), 6895-6906. https://doi.org/10.1007/s12652-021-03589-9

[8]. B. M. Rahal, A. Santos and M. Nogueira, "A Two-Tier Distributed Architecture for Effective and Efficient Detection and Prediction of DDoS Attacks," in IEEE Access, vol. 8, pp. 89376- 89387, 2020, doi: 10.1109/ACCESS.2020.2993285.

[9]. A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet, "More efficiency in multiple kernel learning," in Proceedings of the 24th International Conference on Machine Learning (ICML), July 2007, pp. 775-782, edited by S. Sonnenburg, G. Rätsch, C. Schäfer, and B. Schölkopf.

[10]. M. Chen, Z. Xu, K. Q. Weinberger, and F. Sha, "Marginalized denoising autoencoders for domain adaptation," in Proc. 30th Int. Conf. on International Conference, 2017.

[11]. Seo, S., Han, S., Park, J., Shim, S., Ryu, H.- E., Cho, B., & Lee, S. (2021). A Distributed Architecture for Early Detection of Distributed Denial of Service Attacks Using Unsupervised Statistical Learning. IEEE Transactions on Network and Service Management, 18(2), 836- 849. DOI: 10.1109/TNSM.2021.3060827

[12]. Wei, Y., Jang-Jaccard, J., Sabrina, F., Singh, A., Xu, W., & Camtepe, S. (2021). AE-MLP: A Mixture Approach for Effective Identification and Categorisation of Distributed Denial of Service Attacks. IEEE Transactions on Network and Service Management, 18(1), 511-524. doi: 10.1109/TNSM.2021.3065854

[13]. Dimolianis, M., Pavlidis, A., & Maglaris, V. (2021). A signature-based defense strategy against DDoS attacks using XDP middleboxes and supervised machine learning. IEEE Access, 9, 124068-124079. DOI: 10.1109/ACCESS.2021.3104233

[14]. Alamri, H. A., & Thayananthan, V. (2020). A Software-Defined Networking (SDN) DDoS mitigation strategy with a bandwidth management mechanism and an Extreme Gradient Boosting (XGBoost) algorithm. IEEE Access, 8, 88912-88923. https://doi.org/10.1109/ACCESS.2020.2991045

[15]. Trejo, L.A., Argüelles, J.C., Maldonado, M.A. et al. (2019). DNS-ADVP: A Platform for Detecting Ongoing DNS DDoS Attacks using One-Class Categorization and Visual Analytics. Journal of Network and Systems Management, 27, 559–579. https://doi.org/10.1007/s10922-018-9477-2 [13]N. Z.

[16]. UNWS-np dataset, GitHub repository, https://github.com/adeelz92/UNSW- NB15_Python_Visualization. Accessed on: May 12, 2023.

[17]. D. Gavrilis and E. Dermatas, "Real-time detection of distributed denial-of-service attacks using RBF networks and statistical features," Comput., doi: 10.1109/MC.2007.78.

[18]. M. Sharifi, A. Jalali, and M. Rezaei, "Enhancing network security by effectively identifying and categorizing DDoS attacks using machine learning methods," Journal of Network and Computer Applications, vol. 149, pp. 1-12, Jan. 2020. DOI: 10.1016/j.jnca.2019.102436.

[19]. S. Guo and H. Tracey, "Discriminant analysis for radar signal classification," in IEEE Transactions on Aerospace and Electronic Systems, vol. 49, no. 3, pp. 1713-1723, Aug. 2013.

[20]. C. Forbes, M. Evans, N. Hastings, and B. Peacock, "Statistical Distributions," Hoboken, NJ, USA: Wiley, 2011.