UtilitasMathematica

ISSN 0315-3681 Volume 122, 2025

Semi-Supervised Machine Learning Approach for DDOS Detection

Dangenti Keerthi

M.Tech Scholor, Department of CSE, Kakinada Institute of Technology & Science, Divili, India.

Email: keerthibangaru1111@gmail.com

Dr D Mohan Reddy

Professor and Principal, Kakinada Institute of Technology & Science, Divili, India.

***Abstract*- With the exponential increase in data transmission across computer networks in today's world, identifying and preventing dangerous network use has become a paramount issue for network managers and users alike. The high volume of incoming data inundating the target server in the network, originating from several sources and resulting on server crashes or severe slowdowns, poses a significant challenge in distinguishing between malicious traffic from attackers and legal traffic from users. The reason for this is that the excessive amount of network traffic overwhelms the server, resulting in its failure or significantly reduced performance. Therefore, it is impossible to stop the attack by concentrating on a solitary origin. Denial-of-service (DDOS) attacks may be executed by skilled hackers for cyberwar or financial motives, or by internal users as a casual pastime. A primary issue for security administrators is the potential occurrence of DDoS flooding assaults. In previous studies, researchers have attempted to identify flooding assaults by using approaches that rely on both traits and abnormalities. They faced challenges in determining the characteristics of the assault flow. Furthermore, there is a lack of coordination among the constituent nodes of the cohesive network. Because the address that was widely used was created unlawfully, it was difficult to ascertain the origin of the attack. The assault was of brief duration, leaving just a limited window for a countermeasure. Hence, it is essential to create an anomaly detection system to detect and mitigate distributed denial of service attacks (DDOS) while safeguarding the data of legitimate users. The main objective of this thesis is to devise a method for protecting stored data by categorizing arriving packets and making decisions based on the categorization in line with the outcomes of the categorization process.**

To detect distributed denial of service attacks in networks, the suggested intrusion detection solution utilizes the optimum weight of the DLNN. The suggested system utilizes the NSL-Data Set for network intrusion detection. This dataset is appropriate for identifying network intrusions and includes additional attributes. The dataset contains a total of 41 distinct categories of training and testing data, enabling it to accurately identify network assaults.The method is initiated by the training phase, which is then followed by the testing phase. The NSL-Data is collected and undergoes preprocessing before being inputted into the neurons of a Deep Learning Neural Network (DLNN). Preprocessing includes methods like as data normalization, missing value imputation, and the conversion of strings in data sets to numerical values.

Keywords: **YOLOv4, Kalman filter, framework, and Youtube.**

I. INTRODUCTION

Network security encompasses any measures used by an organization to thwart unwanted access or inadvertent harm to user devices, sensitive data, or the network infrastructure. The primary objective of network security is to provide uninterrupted access to the network for authorized users while safeguarding its operational integrity. Reacting promptly to security threats is a crucial component of ensuring the security of an organization. An essential stage in the process of addressing a security problem is to determine the veracity of an occurrence. Identifying vulnerabilities may be achieved via several approaches, such as using an intrusion detection system, conducting detection studies, and gathering information from end users and other stakeholders inside the organization. Intrusion detection, often referred to as ID, is a complex task that requires extensive expertise in security and a comprehensive understanding of both the system and the company by security professionals. The growing need for constant access to personal communications has driven the creation of innovative networking methods. Information security is a crucial aspect of data transmission, and its significance has increased due to the widespread use of networks by individuals. Various methods are used to enhance the security of transmitted data. This chapter comprises an exposition of the contextual information, an analysis of the topic, research obstacles, objectives, and the structure of the paper.

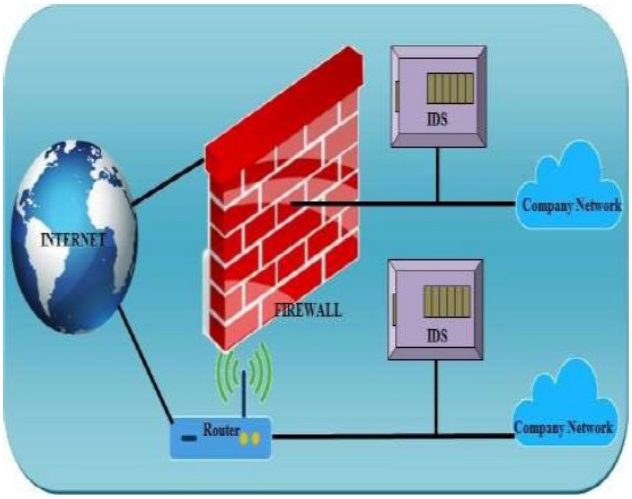

Figure 1: Intrusion Detection System

A host-based intrusion detection system (IDS) is a computer infrastructure monitoring system that analyzes network traffic and records any harmful behavior. In

addition, Host-based Intrusion Detection Systems (HIDS) provide comprehensive surveillance by effectively monitoring the critical security aspects of the systems.

II LITERARURE SURVEY

Establishing authentication methods for certain users and devices, protecting data, ensuring compliance with legislation, and maintaining consumer privacy are all functions that are accomplished via the implementation of network security measures. This is because there is a growing need for

The proliferation of network systems has resulted in a considerable increase in the level of concern about network security among internet users as well as among internet professionals. Therefore, it is of the utmost importance to safeguard the data on the network by suggesting an intrusion detection system that is very efficient.

According to Liao et al. (2013), an intrusion detection system is a tool that assists in the identification of illegal alterations that are carried out by malevolent intruders in computer system files. The Distributed Denial of Service (DDoS) assault is the most common sort of attack that may be detected by an intrusion detection system that is installed inside a network. But this system can also detect other types of attacks. Within the context of the network system, this specific assault has the potential to damage both the system's security and its dependability. It is possible that these assaults may result in server-based problems such as privilege violations and unauthorized logins, as well as vulnerabilities in network services and the manipulation of data by apps. Through the process of monitoring system activity and categorizing it as either normal or strange, an anomaly-based intrusion detection system has the capability of detecting computer and network invasions, in addition to abuse. The selection of the right classification and detection strategy is the first obstacle that must be overcome in the process of anomaly detection. This is done in order to reduce the number of false positives and improve accuracy.

Anomaly-based intrusion detection systems (IDS) that make use of intelligent classification techniques have been shown to be more successful in spotting denial-of-service (DDOS) assaults in comparison to other ways. This is despite the fact that several methods have been presented in the recent literature on network security. This chapter presents a detailed review of the research that has been conducted on anomaly-based intrusion detection systems for denial-of-service attacks, including comments that are pertinent to the topic.

In the context of a network, an anomaly is a divergence from the defined attack methods or an anomalous behavior that is demonstrated by the network. The number of instances in which false positive alarms occur is substantial. A capability that enables one to spot noteworthy patterns is referred to as anomaly detection. The ever-increasing reliance on technological platforms has resulted in a rise in the amount of information on network traffic that is being sent. Because they often provide the most accurate picture of irregularities, visualizations are quite useful when it comes to the process of developing and testing models for anomaly detection applications. An analysis is performed on the following list of approaches that are often used to discover anomalies: The phrase "Ismail et al., 2008" is a reference to a research that was carried out in the year 2008 by Ismail and his colleagues. In the context of congested surroundings, a novel approach was proposed for the detection of abnormalities on a global scale. By using the optical flow of frames, it was possible to separate layers of individuals who were moving through the crowd from the foreground.

Through the use of the optical flow calculation that occurs between two successive frames, a single layer is produced. In order to improve the effectiveness of the extracted layers, a reliable method known as the artificial bacterial colony meta-heuristic is used. For the purpose of covering all places with considerable frame movement, artificial microorganisms are often used. The artificial bacterial colony is able to quickly adapt to a broad variety of different settings. In addition, the method is very resistant to noise as well as rapid shifts in the lighting of the video, as shown by the optical flow data. The training of a Kohonen's neural network included the use of inputs such as the location of the colony's centroid in respect to each optical flow layer, the colony's capacity for food storage, and the number of bacteria that it contained. The network revealed the capacity to distinguish between certain episodes by analyzing the behavioral patterns shown by the bacterial colony during various events. This ability was proven when the network was given instruction.

In their study, Chakraborty et al. (2013) revealed that the UNSW-NB15 data set exhibited three separate features of complexity. Initially, a detailed explanation of the statistical analysis of the data and characteristics was presented to the audience. Additionally, a feature correlation analysis was presented for your consideration. A further evaluation of the complexity was carried out by using the accuracy and false alarm rates (FARs) of five different classifiers that were already in existence. Next, the KDD99 dataset was used in order to do a comparison between the outcomes. It has been determined, on the basis of the findings of the experiments, that the benchmark data set known as UNSW-NB15 is preferable than KDD99 when it comes to assessing Network Intrusion Detection Systems (NIDS). Compared to KDD99, the complexity of UNSW-NB15 is much higher.

Researchers came up with a fuzzy-based technique (Shanmugavadivu et al., 2011) for spotting intrusions in order to handle the increasing number of people who utilize networks. When it came to identifying intrusions, this system shown a high degree of consistency and accuracy.
According to Farid et al. (2015), the network attack was discovered via the use of a number of different

techniques, including the Hidden Markov Model, the Self-Organizing Map, and the Decision Tree models. A comparison and recording of the outcomes of various models was carried out. The metrics, which include the rate of identifying threats, the degree of accuracy, the amount of time necessary for training, and the proportion of false positives, were compared to the data that was supplied in the tables for these particular characteristics.

In order to develop a model that is flexible enough to accommodate the typical daily patterns of network traffic, the strategy that was shown by Luigi et al. (2013) depends on the use of many data mining techniques. There are four primary phases involved in this procedure. The usual model was constructed on the basis of a specific collection of symbolic objects, whilst the initial observation units, such as the links between the networks, were translated into symbolic objects. In the event that a newly added symbolic item differs from the things that are already present in the model, this change is regarded as an anomaly. It is possible that it will be included into the model if it is determined to be a point of change. Unless otherwise specified, it is considered to be an anomaly. It is possible for a network administrator to utilize the network connection model that was built in order to spot inconsistencies in the patterns of network traffic that may need further investigation.

An examination of the firewall logs was carried out using the method that was requested. For their introduction research in 2013, Guising and colleagues carried out an exhaustive investigation of the many possible methods that may be implemented to address the problem of network anomaly detection. In contrast to signature-based detection methods, data mining has the capacity to automatically identify typical prototypes from enormous volumes of network data and separate them from one another. This is far more effective than signature-based detection methods. However, because of the intrinsic properties of network data and the complexity of the procedures involved, it is not possible to simply apply data mining techniques to solve the issue at hand. Some examples of these approaches are feature selection, clustering, association rules, and classification. They suggested a number of changes after conducting an exhaustive investigation into the deficiency, with the goal of detecting irregularities in a timely and accurate manner.

Statistical analysis tools for network data are becoming an increasingly important component in the area of cybersecurity technology. In a business network, the amount of data and the rate at which it is sent Streaming analytics are highly valued by sources because of their capacity to analyze data just once while simultaneously successfully controlling temporal volatility. Identifying abnormalities in real-time networks is a method that was presented by Jordan et al. (2017). The method starts by locating any irregularities that may be present in the correlation processes that are taking place on each edge of the network graph. Additionally, deviations from a large number of boundaries were meticulously documented and assessed in order to improve comprehension of the overall state of the network. An example of the technique was shown and evaluated via the use of two genuine Net flow datasets in a simulation. The K-Nearest Neighbor strategy was shown to increase the cross-digestion rate and performance, as well as the precision, duration, and consumption accuracy, according to the results of a detection experiment that was carried out by Fengchen et al. (2018).

The K-Medoid clustering and K-Nearest Neighbor classification algorithms were used by Yu et al. (2018) in order to identify intrusions in huge datasets. This model's performance was tested by calculating the accuracy and the reliability of the results

Detecting distributed denial of service attacks in networks was accomplished by the authors via the use of K-medoid clustering and the K-Nearest Neighbor technique (Jin et al., 2018). This resulted in an improvement in cyber security and the protection of vital infrastructure.

Hasan et al. presented the Artificial Neural Network (ANN) model in 2019 with the purpose of identifying risks to networks that had not yet been detected. A Distributed Denial of Service (DDoS) assault was launched against the network, and the results were effectively displayed by the ANN model for DDoS detection. A thorough analysis was performed to determine the model's accuracy, sensitivity, and specificity.

III MATERIALS AND METHODS

The internet is a global computer network that is linked to one another via a variety of different media and functions according to a standard protocol. The Internet is an essential component in the lives of modern people since it serves a multitude of functions, including the facilitation of economic transactions, the facilitation of social connections, the provision of educational resources, and the provision of entertainment option choices.

There are many significant facets of human existence. Within the realms of communication and computers, the Internet is often considered to be the greatest important technological innovation that has ever occurred. Theft of network resources, harm to one's reputation and brand, financial damages, identity theft, loss of private information or data, theft of network resources, and a drop in customer trust in online banking and e-commerce are only some of the hazards that may be incurred as a result of web threats. As a result of the separation of data and business logic on a distant network server that is not subject to visible management, the majority of security concerns are distinct from those that were previously present in infrastructures that did not include networks. The denial of service attack, often known as a DDoS attack, is a kind of cyber attack that has been exhibiting behavior that is becoming more aggressive and alarmingly intrusive against internet services. The server or collection of computers that makes a service available to

its customers is often the target of distributed denial of service attacks. According to Parivindar et al. (2016), the objective of denial-of-service attacks is to overwhelm a server that is working by flooding it with an excessive amount of requests. This will cause the service queue to become overloaded, which will result in the client being unable to access the service.

Distributed denial of service assaults, also known as DDoS attacks, are able to be launched against hacked systems that are used by residential buildings, educational institutions, and government entities. Bots are a common term used to refer to these types of systems. Denial-of-service assaults, often known as DoS attacks, are typically launched at the network layer.

Packages of ICMP, SYN, or UDP that are excessive. During an application layer distributed denial of service attack, the attacker shifts their attention to the application layer and overwhelms the system with a huge number of HTTP GET packets. This occurs after the attacker has been unsuccessful in their attack on the network layer. It has been stated by Amit et al. (2018) that distributed denial of service attacks may be carried out via a variety of techniques, including TCP SYN, UDP flood, DNS reflection, HTTP flood, and ICMP flood. Information security research conducted in recent years has shown that governments and businesses all around the globe have suffered large financial losses as a result of distributed denial-of-service assaults, often known as DDoS attacks.

By using the computing capacity and geographic dispersion given by a large variety of devices and their different movement patterns, attackers are able to utilize more sophisticated tactics to increase assaults and overwhelm victims. This is often accomplished within the context of an Internet of Things (IoT) network scenario. As a result, it is essential to have a DDoS detection solution that is not only effective but also practical.

The network layer of an Internet of Things (IoT) system is often where an intrusion detection system (IDS) operates in order to improve the system's level of information security. An Intrusion Detection System (IDS) for an Internet of Things (IoT) system has to be able to analyze data packets and deliver instant replies in order to be considered effective. Using several protocol stacks, it should be able to scan data packets across different layers of the Internet of Things network. Additionally, it should be able to adapt to a wide variety of Internet of Things environment technologies. The intrusion detection systems (IDS) are developed specifically for Internet of Things (IoT)-based smart environments, which need them to function in difficult settings with limited computing resources, short reaction times, and intensive data processing requirements. In light of this, it is possible that traditional intrusion detection systems (IDSs) are not wholly appropriate for contexts that use the Internet of Things (IoT). The problem of Internet of Things (IoT) security is a persistent and significant one, which highlights the need of maintaining a current knowledge of the most recent information on vulnerabilities in IOT system security and devising effective ways to address these vulnerabilities (Faisal et al., 2020). When it comes to assuring the safe storage of client information in network-based Internet of Things systems, security is of the highest significance. As a result, network security is not only important for individual users but also for corporate users. Businesses are required by law to protect the data of their customers, and many industries have rules that are more rigorous than others regarding the preservation of data. It is essential for individuals to have the confidence that their personal information is safeguarded and free from any possible breaches or threats.

Network computing security involves a wide range of difficulties, including multi-tenancy, data loss and leakage, network accessibility, identity management, hazardous application programming interfaces (APIs), inconsistent service level agreements, patch management, and internal threats. According to Varun et al.'s research from 2020, traditional basic cryptographic algorithms are lacking in a number of security characteristics because of their inadequacy, incompatibility, and low scalability. " In order to satisfy the need for network security against distributed denial of service attacks (DDOS), a solution has been developed. Through the implementation of an Intrusion Detection System (IDS) that integrates an encryption strategy, this solution may be effectively implemented.

A. Proposed Model Of Ids For Ddos Attack Detection

For the purpose of effectively identifying both normal and attacked data in a network, a complex model that is developed using DLNN combines optimal weights in the hidden layers. The detection of distributed denial of service attacks in networks is the primary purpose of this approach. The explanation about the

For your convenience, the suggested prototype is shown below.

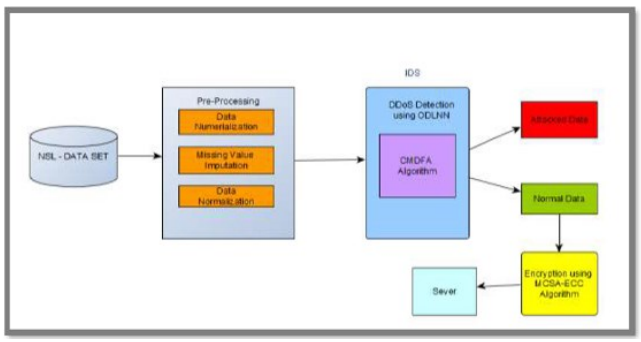

Figure 3.1: Proposed Model to detect DdoS Attack.

The DDOS volumetric attack is one of the most dangerous forms of malicious traffic on the internet. To incapacitate the victim's system, several attackers collaborate to distribute a substantial amount of meaningless data in this volumetric attack.

Computer resources or proximity to network

connections. The proposed safe encryption approach is based on MCSAECC, while the intrusion detection system for detecting DDOS attacks is based on ODLNN. This technique involves both training and testing phases. During the training phase, the NSL-KDD dataset undergoes preprocessing which involves tasks such as data normalization, data nominalization, and replacing missing attributes. To distinguish and classify the data as either normal or attacked, it is advisable to use the approved classification method during the training phase. The subsequent steps in the process include preprocessing and classification, similar to what was done during the training phase. Should the classified data be deemed ordinary, it may be subjected to encryption using the MCSA-ECC method and stored on a network to enhance protection against unauthorized access. In the future, if we need access to encrypted authentic data, we will decipher and use the networked data. If the compromised data is not removed, it is retained on the network as a log file for the purpose of identifying future attacks. The NSL-Data collection consists of 41 distinct features, categorized into two groups: normal and attacked.

IV RESULTS

The readout from the sensor is at 2000. The artificial neural network (ANN) that is currently in use provides a result of 92.66, but the optimum deep learning neural network (ODLNN) that is suggested generates a value of 94.66. With the number of nodes set at 3000, 4000, and 5000, respectively, the current artificial neural network (ANN) approach generates results of 93.44, 95.34, and 96.46. These are the levels of accuracy achieved by the technique. The accuracy rates of 94.67%, 96.34%, and 97.46% are produced by the suggested method that is based on ODLNN. These rates are shown in sequential order. When compared to the approach that is currently being used, this demonstrates that the strategy that is being advocated produces far better accuracy figures. In Figure 4.1, a graphical depiction of the accuracy values is shown for your evaluation.

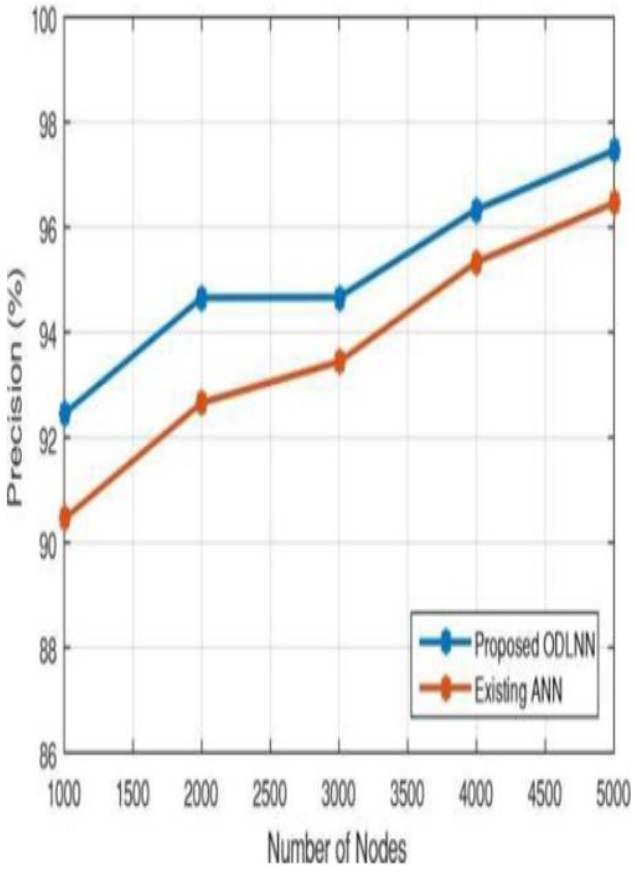

Figure 4.1:Graphical Representation precision values

Table 4.2:Comparison of Recall value between Exsting ANN and proposed ODLAN

Number of Nodes	Existing ANN (%)	Proposed ODLNN (%)
1000	92.12	93.11
2000	91.89	93.35
3000	92.23	93.23
4000	93.89	94.83
5000	95.89	96.83

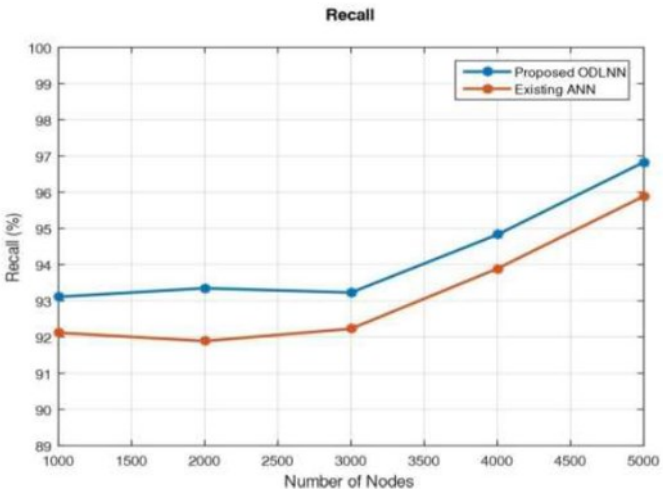

Figure 4.2: Graphical Representation recall values

Table 4.3:Comparison of F-score values between Exsting ANN and proposed ODLAN

Number of nodes	Existing ANN (%)	Proposed ODLNN (%)
1000	90.23	92.05
2000	91.11	93.15
3000	92.23	93.89
4000	93.57	95.14
5000	95.89	96.85

The values of the F-score metric that were obtained for a variety of Number of Nodes are shown in Table 4.3. According to the findings, the current Artificial Neural Network (ANN) obtains an accuracy of 90.23% for a dataset consisting of 1000 nodes, but the newly developed Optimized Deep Learning Neural Network (ODLNN) achieves an accuracy of 92.05%. Concerning the sensor in question

The value of 2000 is connected to an ODLNN value that is indicated to be 93.15 and an ANN value that is now present to be 91.11. The results that are produced by the current artificial neural network (ANN) technique are 92.23, 93.57, and 95.89 for the Number of Nodes of 3000, 4000, and 5000, respectively.

Accuracy ratings of 93.89%, 95.14%, and 96.85% correspondingly are produced by the use of the ODLNN-based approach that has been suggested. When compared to the approach that is currently being used, the F-score values that were obtained for the proposed strategy are much higher. Figure 4.3 is a graphical depiction of the data obtained using the F-measure algorithm.

CONCLUSION

Intrusion detection is a method of network security that is used to identify, prevent, and prohibit illegal access into networking systems that are used for communication or computer networks. The use of Intrusion Detection Systems, often known as IDS, is very necessary in order to keep a system's security and integrity intact.A network that is very safe and unaffected by any potential dangers. In the context of network intrusion detection, abnormality-based intrusion detection refers to a collection of methods that are designed to categorize network data as either typical or unique. Denial of service assaults, also known as DDoS attacks, are a complex and extensive issue that arises when computers deliberately target and disrupt the infrastructure of a network, hence inflicting serious damage to the system of a person or an organization. Attacks that are classified as Distributed Denial of Service (DDoS) also have an impact on the quality of the service. To put it another way, even if the user's data transmission rate would drop, the time delay will also rise. DDoS assaults are difficult to detect, which is a serious concern. Despite the fact that the attacker is the one who starts the attack via the compromised system, they do not take part in the assaults themselves. Consequently, it is difficult to ascertain the source of the assaults given the circumstances.
One of the primary goals of a denial-of-service attack is to identify the Internet Protocol (IP) addresses of the routers that the packets are traveling through. There are three different analytical approaches that are used in order to efficiently protect data while simultaneously identifying threats. These approaches are as follows:

Using a deep learning neural network, the unattributed distributed denial of service attack was identified. For the purpose of fine-tuning the weights belonging to the DLNN, the suggested technique makes use of CMDFA. After the data was discovered to have been subjected to an attack, it was stored on the network and encrypted using MCSA-ECC in order to further strengthen the data security measures. Following that, the apps were granted permission to retrieve the encrypted data in order to proceed with the operational procedures. The findings of the secure data encryption (MCSAECC) and attack detection (ODLNN) mechanisms were compared to those of the approaches that were already in use in order to determine whether or not the suggested methods were effective. Both the ODLNN and the current ANN were evaluated in terms of their accuracy, f-measure, precision, and recall. A comparison was done between the two. A comparison between the Artificial Neural Network (ANN) and the Deep Learning Neural Network (DLNN) reveals that the DLNN consistently generates the highest values for each and every statistical measure. This demonstrates that the DLNN is more effective than the ANN. A comparison was also made between the MCSA-ECC and the existing ECC with regard to the amount of memory that was used, the speed at which encryption and decryption occurred, and the degree of security.

REFERENCES

[1] Abdulkadir Abdullahi Ibrahim and Wilson Cheruiyot Michael. (2017). Data Security in Network Computing with Elliptic Curve Cryptography.*International Journal of Computer (IJC).*, 26(1):1-14.

[2] Ahmad Riza'ain Yusof., Nur Izura Udzir., and Ali Selamat, (2019). Systematic literature review and taxonomy for DDoS attack detection and prediction.

[3] International Journal of Digital Enterprise Technology., 1(3):292-315.Akira Yamada, Yutaka Miyake, Keisuke Takemori, Ahren Studer, and Adrian Perrig. "Intrusion detection for encrypted web accesses". *Proc. 21st*

[4] *International Conference on Advanced Information Networking and Applications Workshops (AINAW'07) ., IEEE,* pp:569-576.Alireza Askarzadeh (2016). A novel metaheuristic method for solving constrained engineering optimization problems: crow search algorithm.Computers and Structures. *An International Journal of Computer and Structure.,* 169(9): 1-12.

[5] Apurv Verma and D. Kumar Xaxa, (2016). A survey on HTTP flooding attack detection and mitigating methodologies. *International Journal of Innovations and Advancement in Computer Science,* 5(5): 18-21.

[6] Aqeel Sahi., David Lai., Yan Li., and Mohammed Diykh (2017). An efficient DDoS TCP flood attack detection and prevention system in a Network environment. *IEEE Access,* 5:6036-6048.

[7] Archana Patel KM, and Prateek Thakral. (2016) . The best clustering algorithms in data mining, proc,

International Conference on Communication and Signal Processing (ICCSP), IEEE, 2042-2046, 2016.

[8] Arun Raj Kumar P., and S. Selvakumar., (2011). Distributed denial of service attack detection using an ensemble of neural classifier. *International Journal of Computer Communications.,* 34(11):1328-1341.

[9] Ashish Kumar Khare, J. L. Rana, and R. C. Jain. (2017). Detection of wormhole, blackhole and DDOS attack in MANET using trust estimation under fuzzy logic methodology. *International Journal of Computer Network and Information Security.,* 9(7):29.

[10] Bharanidharan, R., and Santhosh R., (2019). Group hash function-based enhancing network security for network service providence. *Journal of Soft Computing.,* 23(18):8495-8502

[11] Bin Jia, Xiaohong Huang, Rujun Liu, and Yan Ma.(2017). A DDoS attack detection method based on hybrid heterogeneous multi classifier ensemble learning. *Journal of Electrical and Computer Engineering.* 28(12): 124-135

[12] Bin Li, Qinglei Zhou, Xueming Si, and Jinhua Fum (2018). Mimic encryption system for network security. *IEEE Access,* 6:50468-50487.