
Kamaljeet Kaur¹, Parveen Kakkar ²

¹ P.G. Student, Department of Computer Science & Engineering, DAV Institute of Engineering and Technology, Jalandhar, Punjab, India
² Associate Professor, Department of Computer Science & Engineering, DAV Institute of Engineering and Technology, Jalandhar, Punjab, India

ABSTRACT: The key objective of distributed denial of service attack is to compile the multiple systems across the internet with infected agents and these agents are designed to and programmed to launch the packet flood. With the increase in popularity of internet there are number of security issues and to handle these issues intrusion detection system (IDS) and intrusion prevention systems (IPS) are employed. IDS and IPS system follows the two different approaches for detecting intruders: Signature based detection and anomaly based detection. Signature detection technique is a method of searching the network traffic for a series of bytes or packet and then compares these packets against a set of signatures from known malicious threats. The anomaly based detection technique is a concept of a baseline for the network behaviour. Baseline can be considered as description of the type of network behaviour that can be accepted, any deviation from this baseline is considered as an anomaly. Therefore anomaly based intrusion detection uses machine learning techniques to detect whether a packet is intrusive or non-intrusive. This paper provides a systematic review of machine learning techniques used in DDoS attack detection.


I. INTRODUCTION

With the internet increasing popularity, they encounter number of security issues these days. Distributed denial of service attack is one of the significant threats among all security issues. DDoS attack is launched by sending a large amount of traffic to target system. They can damage one or a group of devices and their resources. Attackers scan the network to find the machine having vulnerability and these machines are used as the agents by the attacker. These are called the zombie. DDoS attacks are initiated by the zombies which are well structured and widely dispersed nodes. Attacker launches the attack with the help of zombies. These are the secondary victims.
DDoS attack comprises of masters, handlers, agents and the victim/attacked server. Zombies are the one used by the master to form a botnet. Masters communicate with agents via handlers. Handlers can be programs installed on a set of network servers. Attackers send command and control their agents through handlers.[4]

A. Classification of DDoS attack

![Fig2. Taxonomy of DDoS attacks](image)

Bandwidth depletion attack: This type of attack consumes the bandwidth of the target system by sending unwanted traffic to the system. Tool like Trinoo is used to perform this type of attack.

Flood attack: This attack sends the huge amount of traffic with the help of zombies that clogs up the network bandwidth with IP traffic.

Amplification attack: In this attacker sends the large number of packets to a broadcast IP address. Fraggle and smurf are the types of amplification attack.

Resource depletion: this attack targeted to exhaust the victim system resources so that the legitimate users are not serviced.

Protocol exploit attack: This attack consumes the surplus quantity of resources from the targeted system by exploiting the specific feature of the protocol.

Malformed packet attacks: In this the packet wrapped with malicious information and attacker sends this packet to crash the target system. [4]

B. Types of DDOS attack

Some specific and particularly popular and dangerous types of DDoS attacks include:

UDP flood attack: An UDP flood attack is initiated by sending a large number of UDP packets to random ports on a remote host. On receiving the packets, target system looks the destination ports to identify the applications waiting on the port. When there is no application, it generates an ICMP packet with a message “destination unreachable”. This process saps host resources and can ultimately lead to inaccessibility.

ICMP flood: Similar in principle to the UDP flood attack, an ICMP flood overflows the target resource with ICMP Echo Request packets, generally sending packets as fast as possible without waiting for the replies. This type of attack can consume both outgoing and incoming bandwidth, since the target server will attempt to respond with ICMP Echo reply packets, resulting a significant overall system slowdown.

SYN flood: In a SYN flood attack, the attacker sends several packets but does not send the ACK back to the server. The connections are hence half opened and consuming server resources. A legitimate user, tries to connect but the server refuses to open a connection resulting in denial of service.

Ping of death (POD): it makes victim host unavailable by sending it oversized ICMP packets as ping request.
Slowloris: Slowloris is a highly targeted attack, enabling one web server to take down another server, without affecting other services on the target network. It accomplishes this by creating connections to the target server, but sending only a partial request.

II. RELATED WORK

Various methodologies and techniques for reducing the effects of DDoS attack in different network environments have been proposed and evaluated. Feng qiaojuan and Wei Xinmeng[11] proposes an improved network security model and analyzed its advantages. They combined the features of current and proposed security model. Jin li yong liu[2] presented a neural network for ddos attack detection to analyse the server resources and network traffic and they use the learning vector quantisation neural network for post anomaly detection. Ruiping luo and lin chang yow[30] introduced intelligent fast flux swarm network and used the intelligent water drop algorithm for distributed and parallel optimization and fast flux used to connect the swarm nodes, clients, servers. They show that simulation consisting of 400,000 clients node and 10,000 swarm nodes can maintain 99.96 percent packet delivery ratio when the network is under attack. Akilandeswari and shalinie[3] have introduced a probabilistic neural network based attack traffic classification to detect different DDoS attacks. The authors focus on separating flash crowd event from denial of service attacks. They have used the Bayes decision rule for Bayes inference coupled with Radial basis function neural network for classifying DDoS attack traffic and normal traffic. Bansidhar Joshi[13] tested the efficiency of a cloud traceback model dealing with DDoS attacks using backpropagation neural network. Cloud traceback check by using the flexible deterministic packet marking which provides a defense system for finding out the real sources of attacking packets. Mohamed Karim and Belhassen[8] provided a description of massive DDoS attacks and a conceptual framework for detecting and reacting to this type of attack in coordinated fashion. This approach proposes to enhance the security level of SAHER's architecture. The process continues in three rounds: detection phase, exchange of intrusion data, response phase. Mehdi Bharti[1] used the new hybrid detection method using genetic and artificial neural network and deployed for feature selection and attack detection and the multi-layer perceptron of artificial neural network was used to improve the detection rate and accuracy of the system. Javed Asharf[12] analyse the various machine learning techniques which can be used to handle the issues of intrusion and DDoS attacks to software defined networks. Keisuke and Vitaly[6] analyzed a large number of network communication packets and implemented a DDoS attack detection system using the patterns of each IP address. The author selected the SVM with RBF (gaussian) kernel to train and test the DDoS attack detection system. Yuri G. Dantas[7] proposes the technique for mitigating network layer DDoS attacks. They proposes a defense for application layer DDoS attacks based on adaptive selective verification. The defense system is formalized in computational system Maude and evaluated by using statistical model checker PVeStA used to prevent ADDoS. Qiao yan and Richard[5] discuss the new trends and characteristics of DDoS attacks in cloud computing and provide a survey of defense mechanism against DDoS attack in software define network. Author reviewed about launching DDoS attacks on control layer, infrastructure layer and application layer of SDN and methods against DDoS in SDN. YunheCuia and LianshanYan[9] proposed mechanism consisting of four modules attack detection trigger, attack detection, attack traceback, attack mitigation. Mechanism is evaluated on SDN testbed. Experimental results show that the method can quickly initiate the attack detection. Opeyemi osanaiye[10] presented a taxonomy of different types of cloud DDoS attacks and DDoS defense taxonomy. Author reviewed academic literature on DDoS attacks against cloud services and the mitigation strategies published between January 2010 and December 2015.

III. MACHINE LEARNING TECHNIQUES

Signature based intrusion detection system require administrator to include rules and signatures to detect attacks. It requires several man hours to test, create and deploy these signatures and again create new for unknown attacks. Anomaly based IDS based on machine learning technique provides the solution to this problem they help in generating a system that can learn from data and provide prediction for the unseen data based on the learned data. Following fig shows some of the commonly used machine learning techniques for DDOS attack detection.
Fig3. Machine learning techniques

Machine learning techniques

Artificial neural network
Support vector machine
Fuzzy logic
Decision tree
Genetic algorithm
Navies bayes
K-means clustering

a) Artificial neural network

Artificial Neural networks (ANNs) are a family group of models inspired by biological neural networks which are accustomed to estimate on a big number of inputs and are usually unknown. ANNs include an accumulation of processing elements interconnected together and aimed to transform some inputs to some desired outputs. In this, the Multilayer Perceptions MLP has been widely adopted neural network for intrusion detection in conventional systems. In NN based packets classification system, each component of the feature vector has one input node. Also, usually one output node is employed for each class to which an element might be assigned. The hidden nodes are linked to input nodes and some initial weight assigned to these connections. These weights are adjusted during working out process. Back-propagation rule is among the learning algorithms employed for MLP based ANN. Propagation rule works on a gradient descent method. This approach calculated an error function which will be the difference involving the output calculated by the network and the output desired. The Mean Squared Error (MSE) is employed to define this error function. The MSE is added over the entire training set. To understand successfully, the real output of network must be brought near to the desired output. This is performed by reducing the worthiness of the error, continuously. The error for a certain input is calculated using back-propagation rule and then this error is back-propagated from layer to the last one. The weights of connections involving the nodes are adjusted based on the back-propagated error. In this manner error is reduced and the network learns. The input, output, and hidden layers neurons are variable. Input/output neurons are changed on the basis of input/output vector. Hidden layer neurons are adjusted according to performance requirements. More the hidden layers neurons more complicated could be the MLP.

Jin li yong liu [2] presented a neural network for DDoS attack detection to analyse the server resources and network traffic and they use the learning vector quantisation neural network for post anomaly detection. Akilandeswari and
shalinie [3] have introduced a probabilistic neural network based attack traffic classification to detect different DDoS attacks. The authors focus on separating flash crowd event from denial of service attacks. They have used the Bayes decision rule for Bayes inference coupled with Radial basis function neural network for classifying DDoS attack traffic and normal traffic. Mohamed Karim and Belhassen [8] provided a description of massive DDoS attacks and a conceptual framework for detecting and reacting to this type of attack in coordinated fashion. This approach proposes to enhance the security level of SAHER’s architecture. The process continues in three rounds: detection phase, exchange of intrusion data, response phase. Bansidhar Joshi [15] tested the efficiency of a cloud traceback model dealing with DDoS attacks using backpropagation neural network. Cloud traceback check by using the flexible deterministic packet marking which provides a defense system for finding out the real sources of attacking packets. Javed Asharf [14] analyse the various machine learning techniques which can be used to handle the issues of intrusion and DDoS attacks to software defined networks. The intrusion detection system based on neural network works in three phases:

1) The raw TCP/IP dump data is parsed into form readable by the machine automated parsers.
2) Training of NN is done on different types of attacks as well as on valid data. The input consists of a number of features. The output can assume any one of the two values: normal data or intrusion.
3) Testing is done on the test dataset.

As mentioned earlier, the purpose of Intrusion detection systems based on NN is to classify the normal/valid and attack patterns along with the type of the attack. Thus classification of a single record can be done easily after suitable working out. So, the IDS based on NN can function as an online classifier for the type of attack it was trained for. The NN will be off-line only for small duration when it is gathering information which is required to calculate the features.

b) Support vector machine

Support vector machines are the supervised learning model used for classification and regression analysis. Using this method the given set of examples is marked with one of two categories. SVM training algorithm builds the model that assigns the examples with category. Classification is done by using SVM by extracting attributes from selected training samples. A network connection is selected as a sample. The benchmark dataset like KDD99 will also be used which contains number of network connections attributes captured from various networks. An input space X is defined for each network connection, selecting n attribute characteristics. The vector x (one dimensional) can be used to describe a network connection as under:

X= {x1, x2, ..., xi, ..., xn} where xi, i = 1,2, ... .... ... n, denote the I characteristic value of the sample x. if we have to find only it is a normal or abnormal connection for each network connection, only two states are sufficient to express this. So y = (+1,-1). If y as +1 then it would be classified as normal connection and if y = -1 it would be classified as abnormal connection. T. Subbulakshmi & K. Balakrishnan [25], this paper generates the DDoS detection dataset and detect them using the enhanced support vector machines. Enhanced multi class support vector machines (EMCSVM) is used for detection of attacks into various classes. The performance of the (EMCSVM) is estimated over SVM with various parameter values and kernel functions. Vipin Das [26] conducted an experiment to detect DoS attacks using rough set theory (RST) and support vector machine (SVM). RST was used to pre-process the data. Feature set selected by RST is sent to SVM model to learn and test. These results are compared with principal component analysis (PCA) and shows that RST and SMV increased the accuracy.

c) Fuzzy logic

Fuzzy logic is based on fuzzy set theory which works on reasoning. Techniques based on fuzziness have been used for anomaly detection. The concept of fuzzy logic lets an object to fit in to different classes simultaneously. C Balarengadurai [17] designed a fuzzy logic system to detect exhaustion attacks in IEEE 802.1.5.4 MAC layer. It uses an anomaly based approach and operates in a distributed manner. They strengthened their scheme by two parameters of attack detection namely energy decay rate and attack detection rate. They distinguish attack scenario from the impact of traffic load on network behaviour. R. sharmaugavadi [28] designed a fuzzy logic based system for identifying the intrusion within a network. They use the automated strategy for generation of fuzzy rules, which are obtained from the
definite rules using frequent items. They use the KDD cup 99 dataset for the intrusion detection. Experiments are performed by using this dataset. Stavros n. shields et.al [27] proposed a method for DDoS detection by constructing a fuzzy estimator on the mean packet inter arrival times. This problem was divided into two challenges, first being the actual detection of the DDoS event taking place and second being selecting the offending IP addresses. Author managed to obtain the result under 3sec detection window.

d) Decision tree

Decision tree algorithm is one of the predictive modelling techniques used in data mining, statistics and machine learning for classification. In this a decision tree is used to sort the instances from root node to leaf node. The dataset is learnt and modelled in algorithm therefore whenever a new data item is given for classification it will be classified as learned from the previous dataset. Decision tree algorithm can also be used for detecting DDoS attacks. Hoda Waguih [21], proposed an approach using classification techniques for DDoS detection. They evaluate the performance of J48 decision tree algorithm for detecting DDoS attacks. Compare it with two rule based algorithm, one is OneR and decision table. Dewan Md. [23] proposed a learning algorithm for anomaly based network intrusion detection system that separates attacks from normal behaviours. Identify different types of intrusions using decision tree algorithm. KDD99 dataset is used and their classes are divided into one normal class and four intrusion classes: probe, DOS, U2R, and R2L. Yi-Chi Wu et al.[22], designed a DDoS detection system based on a decision tree technique in which attack is detected then system trace back to the attackers location using a traffic flow pattern matching technique. A C4.5 classifier is used for DDoS attack detection.

e) Genetic algorithm

Genetic algorithm is a process of natural selection that associates with larger class of evolutionary algorithm. GA is a search method that identifies an approximate solution to optimization tasks. GA algorithm based intrusion detection system is used to detect intrusion based on past behaviour. In this method a profile is created for the normal behaviour. Based on this profile GA learns and takes the decision for the unseen patterns. GA also used to develop rules for network intrusion. A typical genetic algorithm requires a genetic representation of solution and a fitness function to evaluate solution [18]. Mehdi Bharti[1] used the new hybrid detection method using genetic and artificial neural network and deployed for feature selection and attack detection and the multi-layer perceptron of artificial neural network was used to improve the detection rate and accuracy of the system. Anurag Andhare [29] generates rules using genetic algorithm based approach to detect DoS attacks. Rule set is generated by training GA on KDD cup 99 dataset to detect attacks. Sing Min Lee and Je Hak Lee[17] proposed an enhance DDoS attacks detection approach which improves traffic matrix building operation and optimizes the parameters of traffic matrix using genetic algorithm. They perform experiments with DARPA 2000 dataset and LBL-PKT-4 dataset for detection accuracy and speed.

f) Navie bayes

Navie bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. It is a family of algorithms based on common principle i.e. value of a particular feature is independent of the value of any other feature, given the class variable. Akilandeswari and shalinie [3] have introduced a probabilistic neural network based attack traffic classification to detect different DDoS attacks. The authors focus on separating flash crowd event from denial of service attacks. They have used the Bayes decision rule for Bayes inference coupled with Radial basis function neural network for classifying DDoS attack traffic and normal traffic. Mouhammd Alkasasbeh [18] compiles new dataset that consist of DoS attacks in different network layers. DDoS attacks are detected using three techniques Multilayer perceptron (MLP), Navie Bayes and Random Forest. MLP showed the highest accuracy rate.
g) K-Means clustering

K-Means clustering is a method of partitioning the dataset into k clusters in which each set belongs to cluster with the nearest mean. Keunsoo Lee & Juhyun Kim [24] proposed a method for proactive detection of DDoS attack by exploiting its architecture. Procedures of DDoS attacks are analysed and select variables based on these features. Then they perform the cluster analysis of proactive detection of attack. The results are evaluated using 2000 DARPA intrusion detection. The result shows that each stage of attack scenario is partitioned well and detects precursors of DDoS attack as well as attack itself. Mangesh D. Salunke[21], proposed an architecture which captures the packets, these packets are manipulated according to requirement such as feature selection, transformation etc. then k-means and naïve bayes techniques are used to classify the packets whether it is normal or is a DoS attack.

IV. PROS AND CONS OF REVIEWED TECHNIQUES

<table>
<thead>
<tr>
<th>S.no</th>
<th>Machine learning techniques</th>
<th>Pros</th>
<th>cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Neural Networks</td>
<td>It can conclude even from limited, noisy, and incomplete data. LVQ and BP neural networks achieve very high recognition rate as compared with others.</td>
<td>The process is not suitable for real time detection because of slow training.</td>
</tr>
<tr>
<td>2.</td>
<td>Support machine vector SVM</td>
<td>SVM used to classify images and achieves high search accuracy. High decision rate and training rate.</td>
<td>Mostly used binary classifier which cannot give additional information about detected type of attack.</td>
</tr>
<tr>
<td>3.</td>
<td>Fuzzy logic</td>
<td>Fuzzy logic is effective especially against probes and port scans.</td>
<td>Depends upon expert knowledge which is not always available.</td>
</tr>
<tr>
<td>4.</td>
<td>Decision tree</td>
<td>Help to determine worst, best and expected values for different scenarios.</td>
<td>Calculations can get very complex when values are uncertain and outcomes are linked</td>
</tr>
<tr>
<td>5.</td>
<td>Genetic algorithm</td>
<td>Genetic algorithm provides best classification rules and select optimal parameters.</td>
<td>Suitability of algorithm dependent upon the knowledge of the problem. Genetic algorithm do not compute well with complexities.</td>
</tr>
<tr>
<td>6.</td>
<td>Navies Bayes</td>
<td>Fast to train and classify and handles real and discrete data.</td>
<td>It may not contain any good classifier if prior knowledge is wrong.</td>
</tr>
<tr>
<td>7.</td>
<td>k- means clustering</td>
<td>Fast, robust and easier to understand gives best result when datasets are distinct from each other.</td>
<td>Applicable only when mean is defined fails for categorical data. Unable to handle noisy data.</td>
</tr>
</tbody>
</table>

V. CONCLUSION

Detecting DDoS attacks using traditional methods is not much efficient. These days machine learning based techniques for detection of DDoS attacks has received much attention. In this paper we have analyzed various machine learning techniques for DDoS detection. The suggested techniques offer an enormous research for both industry as well as academia.

REFERENCES


