International Journal of Engineering & Technology, 7 (2.8) (2018) 472-476

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research Paper

Proposed approach to detect distributed denial of service attacks in software defined network using machine learning algorithms

Shruti Banerjee[1*], Partha Sarathi Chakraborty[2]

[1]M.Tech. CSE, Dept. of CSE, SRM University, NCR Campus, PO Modinagar, Ghaziabad, India
[2]Assistant Professor, Dept. of CSE, SRM University, NCR Campus, PO Modinagar, Ghaziabad, India
**Corresponding Author Email: shrutibanerjee5@gmail.com*

Abstract

SDN (Software Defined Network) is rapidly gaining importance of ‘programmable network’ infrastructure. The SDN architecture separates the Data plane (forwarding devices) and Control plane (controller of the SDN). This makes it easy to deploy new versions to the infrastructure and provides straightforward network virtualization. Distributed Denial-of-Service attack is a major cyber security threat to the SDN. It is equally vulnerable to both data plane and control plane. In this paper, machine learning algorithms such as Naïve Bayesian, KNN, K Means, K-Medoids, Linear Regression, use to classify the incoming traffic as usual or unusual. Above mentioned algorithms are measured using the two metrics: accuracy and detection rate. The best fit algorithm is applied to implement the signature IDS which forms the module 1 of the proposed IDS. Second Module uses open connections to state the exact node which is an attacker and to block that particular IP address by placing it in Access Control List (ACL), thus increasing the processing speed of SDN as a whole.

Keywords: *Software Defined Network; Distributed Deninal-of-Service Attack; Machine Learning Algorithm.*

1. Introduction

A novel and new way of network management is SDN. It was developed as separation of the control layer and forwarding devices which has its own advantages: deploying of new installations and versions are made easy, improves overall network efficiency, straightforward network virtualization management is made easy and it consolidates the middle boxes as a part in the control logic. In SDN, incoming packets are not process the switches; they just forward the packets to its destined node which is identified by matching the incoming packets with the flow table entries. If match is not found for any it is forwarded to the controller for further processing and flow table entries of that request. The controller is the operating system of SDN all the control logic and network functionalities are stored in the controller of SDN. This centralized control logic in SDN makes it vulnerable to DDoS attacks. Both the data plane and control plane can be a target for the DDoS attack.

The attacker places a successful attack by flooding the nodes that are frequently used in the network by studying the flash crowd and access patterns. This exhausts the bandwidth of the network completely. The main aim of DDoS is to make the SDN network unavailable to its genuine users. Firstly, the attacker will compromise a given node of the network these compromised nodes are termed as BOTS or ZOMBIES. Using these Bots, they install all the malicious software for placing a successful attack over the controller of the SDN. If the attacker successfully infiltrates the controller it can gain access to the data and resources of the network which compromises the entire network.

In this paper, IDS is proposed which uses Machine Learning algorithms to classify the traffic as normal or abnormal so as to prevent an intruder to affect the network and its resources. The IDS has two functions: First function uses the signature IDS that implements the trained Machine Learning algorithm for classification of traffic. If it finds any anomalous behavior, it forwards that the hosts set to the next function which has the advanced IDS that checks for open connections and gives the correct result stating which among the hosts set forwarded is the attacker or intruder. This IP address is then entered in the Access Control List (ACL) of the SDN.

2. Literature Survey

Software Defined Network [1] enhances the network management by separating the data plane and control plane. This also makes it vulnerable to various Denial-of-Service attacks. The controller of the SDN allows operations to express high level policies; these policies are made into low-level flow rules installed in the switch. Switches have only one responsibility to forward the packets according to the flow rules. The switches have two modes in which these rules are installed in them i.e. proactive and reactive mode. In proactive mode, the controller breaks all the policies into flow rules and installs them at once in the switches. In reactive mode controller will compute and install a flow rule only when the switches ask for it explicitly. Thus switches working in reactive mode, helps to adapt quickly to changing network dynamics. But it makes switches and controller of SDN vulnerable to DoS attacks. The attacker can specifically compromise a host and flood with short lived spoofed flows thus the large numbers of requests send to the controller by switches. These can cause compromised software components, congestion in channel between the switch and controller, saturation of controller’s resources and flow table overflow. The main aim in [1] is to study more systematically about the DoS threats in SDN. The most concerning DoS attack in SDN is controller resource saturation attack. The better way to mitigate this type of attack is to keep a check in the controller side leaving the switch side untouched. To enforce a fair sharing of resources amongst all the nodes and switches of the network causes to face the challenge that there are quite a large number of switches and hosts for which need to maintain a large number of queues.

The countermeasure for said problem is proposed to be a multi-layer fair queuing (MLFQ); a method of queue management that allows expansion dynamically and aggregation of the queue according to its needs. The basic idea behind is to maintain number of queues in the controller, when there are no attacks. Then dynamically expands the particular queue into multiple if the queue size exceeds the threshold value.

Software Defined Network provides a flexible network management by separating data plane and control plane. This separation leads to communication between control plane and data plane which turns to be a severe bottleneck in cause there is extensive communication between these two entities of the network. In [2], a thorough analysis of Control plane saturation attack is studied which exploits this bottleneck caused under high traffic conditions between the two entities; also how it can be amplified by long forwarding paths in SDN network.

The most widely used SDN architecture is OpenFlow, that provides a standard interface between the switch and the controller. OpenFlow works in reactive mode i.e. on arrival of a new flow it sends a request to the controller stating to create and install the flow rules which tells how to manage the incoming packet. On adversary, this network flexibility due to reactive mode of operation introduces new threats and makes the network vulnerable. The attacker exploits the bottleneck caused due to extensive communication between planes and also the reactive mode of functionality of SDN switch by studying the behaviour of traffic generated, to flood the switch with large number of unique flows, as each network flow will forward a request to the controller. This new inbound flow is high enough causing control plane saturation attack. Till date there is no proposed solution that completely tackles the problem of CPSA attack.

The analytical study shows how long paths of forwarding (a path with maximum number of hops in SDN) in SDN is exploited even with limited power attack, but can amplify the effect of CPSA to a critical level. In [2], shows the analysis on the effects of long forwarding paths against the DoS attacks in SDN controller. It states how a well-informed or an intelligent attacker using only limited power, exploits the long forwarding paths to amplify the attack rate to achieve successful CPSA attack.

Software Defined Networks [3] is a new emerging paradigm for programmable networks. In SDN, switches (forwarding devices) are only responsible for one task i.e. the forwarding of packets, if the packets match with the entries of the flow table entries. The controller (OS of SDN network) performs all the major functionality for the network and grants centralized control over the SDN network. Distributed-Denial-of-Service attack is the biggest threat to cyber security for the SDN network. It attacks the network layer and the application layer of compromised systems in the SDN. When it attacks the network layer it causes bottleneck and when it attacks on the application layer it causes exhaustion of the CPU resources. The attacker sends flood traffic to the targeted nodes of the network, which is an intentional attempt to fully exhaust the CPU resources and the bandwidth to disturb the SDN network, thus making it unavailable for its legitimate users. When a large number of the users are accessing the network (i.e. it is the more frequently used nodes in the network) this traffic is termed as Flash crowd. These flash crowd and the access patterns are studied by the attacker to place a successful attack on this frequently used node. Once it successfully infiltrates this nodes becomes compromised, in which the attacker will install all the malicious software and tools to attempt the attack on the server of the SDN network. These compromised nodes are termed as Bots or Zombies. To place an attack on the SDN network, attacker will use the spoofed IP addresses so that these zombies are unique and avoids the chances of getting detected. On successful infiltration of the attacker to the server, it compromises the data, network resources, access right, flow table rules, firewall in general the whole SDN network.

In [3] an Advanced IDS is designed which is made up of two function. First function is the Signature IDS consisting of trained model, which uses machine learning algorithm to train the trained model of the IDS, so as to classify the incoming packets as normal or abnormal. In case, anomalous nodes are encountered it is forwarded to the second function. Second function consists of the Advanced IDS which checks for open connections and gives the exact result of which node is the attacker so that the particular set of IP addresses are blocked by placing them in the Access Control List (ACL). Module 1 reduces the number of nodes for processing by the second function as it is responsible for the processing speed of SDN which is an important factor.

Distributed Denial-of-Service attack makes both the control plane and data plane vulnerable until now, we encountered only the threat and vulnerabilities on the control plane. But, data plane is equally vulnerable. In [4] separation of the planes in SDN network introduces new network security issues. SDN processes in two modes: proactive and reactive. In proactive mode, controller will preset all the forwarding rules according to the configuration of the networking applications at once. Incoming packets that do not match the forwarding rules are dropped and request is terminated for that IP address. In reactive mode, whenever a table miss is encountered it results in sending a query to the controller. The controller will make decisions of the networking applications keeping in mind the global view of the state of the network; then a network policy is enforced. A typical reactive SDN architecture will use the following steps: sends the packet for a new connection to the switch of SDN, packet is encapsulated as a packet-in message and sent to the controller. This packet-in is processed by SDN application providing network functionality. The decision includes a Flow Mod message; installing a new rule then forwarding the original to its ports. Further, packets fulfilling this flow rule will be handled in fast forwarding hardware. Forwarding tables have limited capabilities with regard to the memory abilities. Switches use CAM also known as Content Addressable Memory to perform table lookup. Binary CAM is limited to a maximum of few 100k entries. An attacker with remote capabilities can easily exhaust the switch memory and can cause a DoS attack. A countermeasure to check the severity of these DoS attacks is use of extended table size. To deal with the threat [4] states countermeasures like specific detection mechanism on data plane, light weighted measure to stop flow attack is a novel way of evaluating the analytic means and simulation.

In [5] machine learning algorithms are used for monitoring and detection of the malicious actions in SDN data plane Statistics and features of the network traffic are generated and a network workload test data is required. The efficiency of the Machine Learning algorithms like Learning Vector Quantization, Self organizing Maps is studied. SDN frees the complexity and static nature of traditional distributed network architecture which is obtained by the system abstraction between control plane and data forwarding plane. Thus, provide an opportunity for vulnerabilities of malicious activities. IDS system is an additional monitoring mechanism besides classical security. In [5] the approach used parameters and statistics from SDN flows and creates tuple that are classified by detection mechanisms. Selecting proper identification of malicious traffic is an important issue for this approach.

These are the selected Machine Learning algorithms used for classification of malicious activities: Self Organizing Maps (SOM), Learning Vector Quantization (LVQ1), Multi-pass Learning Vector Quantization (MLVQ1), Hierarchical Learning Vector (HLVQ1). The results of these above algorithms are studied to provide recommendations on their usage for IDS in SDN environment. The analysis states it is possible to achieve an average value of TPR greater than 94%. Thus, HLVQ1 algorithm is an effective way to improve TPR, when compared to SOM<LVQ1, MLVQ1 for all classes.

As we have [6] decoupled control and data plane, SDN can handle the increasing number of attacks by blocking the network connection at switch level. The important fact is where the challenge lies, to define the set of rules on the controller of SDN to do so (i.e. block the malicious attacks). Historical network attack data is used to identify and block malicious attacks. Limited open source software tools are available to monitor the login attempts which are not efficient to act against chain of attacks. Four Machine Learning algorithms are used i.e. Decision Tree, Bayes Net, Decision Table and Naive Bayes to predict the host by analysing historical data.

The decoupling of SDN architecture provides the advantages of the improved network efficiency (in overall). Secure Shell Brute force attack can be performed to possess serious threat on this SDN architecture. Even if the attack is identified it might not be possible to stop it from making significant damage to the network. The need for specific security rules apart from the traditional firewall is encountered. Malicious users have certain common characteristics which can help to identify and differentiate from other legit users. Various Machine Learning algorithms are used to classify the users to prevent the potential attacks from placing a successful attack, using the historical network attack data. The training data used to train Machine Learning model is the Long tail project [9]. The output of the Machine Learning algorithms is the security rules that are to be implemented on SDN controller to resist access of potential attacker.

The impact of Distributed Denial-of-Service [7] ranges from a simple inconvenience in using a particular server to causing major failure in the targeted device. In [7] a method is proposed to detect the Distributed Denial-of-Service attack from traces of traffic. These traffic traces are used to create multi-dimensional access matrix. Principle Component Analysis (PCA) reduces the attributes used for detection. Machine learning classifiers are Naïve Bayes and K-Nearest Neighbour that used to classify the traffic as usual or unusual. The performance of these classifiers with PCA attributes are analysed based on two metrics: detection rate and false positive rate. The attackers use this threat (i.e. DDoS) to mainly disturb the services provided by the networks thus making normal service unavailable to its legit users. The attacker compromises node of the target network called BOT. It then uses that particular BOT to install all the malicious tools and software so as to place a successful attack on the network. They use the spoofed IP address for two reasons: uniqueness and minimizes the chances to get revealed.

On the basis of the nature of DDoS attacks. They are classified as: end-point attacks and Infrastructure attacks. The DoS attacks that targets the network layer is called NET-DDoS, the DoS attack that targets the application layer is termed as APP-DDoS. Defence for the DDoS attacks can be created in many ways. Development of IDS, firewalls and enhanced router will guard against the attack traffic (i.e. flood traffic). The features of traffic in the low layer cannot effectively differentiate between APP-DDoS and the flash crowd. The proposed method to detect the DoS attack is from the HTTP traces. These traces help to generate the access matrix, attributes are selected and is used in classifier to detect the attackers.

Software Defined Network provides a way of interface between networks and applications. Accurate way of classifying the traffic is an important measure. In [8] the four variants of Neural networks estimator are used to categorize traffic. An important case for the network management is to have high availability and efficiency is the traffic classification. These can be various methods to do so such as Deep Packet Mining Inspection, and using port numbers to determine applications and application layer protocols.

These methods have their own benefits and challenges. The goal of traffic classification methods is to enable controller to distinguish and isolate different application flow, management and programming flow to guarantee QoS. [8] Proposes a framework that determines the application type of flow. This framework uses Machine Learning based trainer to receive the information. Decision Tree algorithm helps in network traffic classification. The algorithms like K-Means gather data for getting trained in learning phase and clustering.

3. Related Work

A. Software Defined Network

The basic principle of SDN architecture is the decoupling of the data plane and control plane and a standard information exchange between the two planes. There are two architectural standards for SDN: For CES and Open Flow. In this paper, we discuss about OpenFlow architecture. In OpenFlow, the forwarding tables are termed as switches and the control logic is termed as controller of SDN. The working of OpenFlow is explained as follows: Once a new packet arrives, the header fields are extracted and matched against the matching fields portions of the flow table entry. If the match is found, switch acts according to the set of instruction of the respective flow table match field entry. If no entry is found, then action taken for that packet is decided by the 'table miss' flow table. This flow table specifies the set of instructions to be performed when no match is found for an incoming packet, which might include passing the packet forward for further processing or drop that packet etc.

The block architecture view of OpenFlow is shown in the figure 1 below:

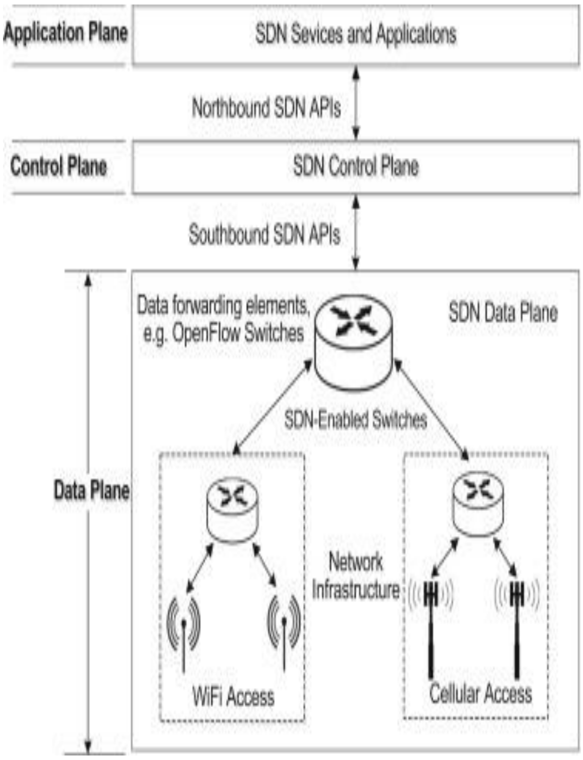

Fig. 1: Block Architecture View of OpenFlow

The forwarding devices are software or hardware devices that perform actions as defined by the flow rules, on incoming packets for eg: drop, forward or rewrite the packets etc. The data plane is the interconnection of these forwarding devices forming a plane termed as data plane. Southbound Interface (SI) is the communication protocol between the switch and controller. Forwarding devices get instructions from southbound API which is a part of SI. The control plane is termed as 'network brain' and all the control logic resides in the control plane. Northbound Interface (NI) is the communication protocol between applications and controller. It takes instructions from SI to program forwarding devices. Management Plane is the set of applications that takes maximum advantage of functions offered by Northbound Interface to implement network control and operational logic.

SDN network architecture has four pillars:

- *The control plane and the data plane* are decoupled. Control functionalities are removed from network devices that will become simple packet forwarding.
- *Forwarding devices* are flow based and not destination based; the flow is defined by the set of packet fields acting as a match (filter) and the set of actions are taken accordingly.
- *Control logic* is moved to an external entity called SDN controller or NOS which provides the essential resources and abstraction to provide the facility of programming the forwarding devices.
- *Network* is made programmable through software applications running on top of the NOS that interact with the underlying data plane.

On the concepts of SDN, It can be classified by three basic abstractions namely Forwarding, Distribution and Specification.

- **Forwarding**: it allows any of the forwarding behaviour desired by network applications hiding the details of underlying network.
- **Distribution**: shield the SDN applications from the distributed state and make the distributed control a centralized unit.
- **Specification:** allows the network applications to express the desired network behaviour without implementing itself.

B. Distributed Denial-of-Service Attack

It is an attack which compromises multiple computer system and attack server or other website or other network resources. These compromised systems (termed as Bots) are used to install all the malware needed so as to place a successful attack on the target (here the controller of SDN).

Distributed Denial-of-Service attack is the abnormal traffic sent to the target (controller of SDN) which is observed as sudden drop of performance in network due to increase in either abnormal delay or CPU utilization or traffic. The attack affects network layer causing Bottleneck and is termed as NET-DDoS. Application-DDoS is known as the attack that targets the application layer causes exhaustion of CPU resources.

The nature of the data and anomalies in it are closely related, they can be studied by various factors such as: packet header information, packet size, delay etc. Flooding the application layer using HTTP GET messages are a part of application DDoS attack. Here such listed attacks - ICMP flood attack, DNS reflection attack, UDP flood attack, and HTTP flood which can helps to achieved DDoS attack.

C. Machine Learning Algorithms

The implementation of the signature IDS using machine learning algorithms are listed as follows:

- *Naïve Bayes:* Naïve Bayes classifier is a conditional probability model based on the Bayes theorem i.e.
-

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

where, P(A|B) is an event A given an event B,

P(A) is probability of A,

P(B|A) is an event B given an event A,

P(B) is the probability of B.

Naïve Bayesian Classifier used for real time prediction, text classification and spam filtering.

- *K-Nearest Neighbour (KNN):* It is a simple form of Machine Learning Algorithm that uses simple algorithm by using similarity measure. It makes prediction using training data set directly. It uses the distance measure such as Euclidean distance measure to identify such K instances in training dataset that are most similar to a new input. If the value of K=1, then it is termed as 1-NN, instant classifier.
- *K-Means:* It is a clustering algorithm that partitions the given number of observations into clusters. Each cluster has been given a mean value. The prediction or classification is done by comparison between the input value and the mean value of the matching clusters.
- *K-Mediods:* It is similar to K-Means but instead of finding mean value for each cluster, each cluster is assigned a middle representative value (termed as mediod). They are used to calculate by using the formula:

$$\sum_{j \epsilon C_i} d(i, j)$$

Here, each object '*i*' is containing by a cluster C_i and the distance between *i* and *j* is measured by *d* (*i*, *j*).

- *Linear Regression:* It is a linear model i.e. a method to model the relationship between scalar dependent variable 'z' and one or more input variables 'w'. If relationship is only a single output variable it is called simple linear regression and for multiple variables used it is called multiple linear regressions. The most common technique is to train the linear regression equation from data is 'ordinary least squares'.

4. Proposed Methodology

The proposed IDS have two functions:

First function is to train the Machine Learning algorithms so as to better predict the incoming traffic as normal or abnormal. The best Machine learning algorithm is used to implement the signature IDS.

Second function uses service three-way handshake to find the exact host which is the intruder and block that IP address by placing it in the ACL.

Using Training dataset to train the Machine Learning algorithms: To obtain accurate trained model so as to identify the potential vulnerable hosts amongst the incoming packets. The main aim of train these ML algorithms are to help identify the attack pattern using the given training dataset. Based on the IP of the attacker, it can predict the potential host that could be an attacker.

These predictions are used to define the security rules in the flow table entries which determine the set of actions to be performed when an incoming packet is matched with the flow table entries. If an IP address is found to be malicious it entered in the access control list (ACL) for SDN.

Implementation of IDS to detect DDoS attack: The IDS is implemented in two function: Function 1 is the Signature IDS which uses the Machine Learning algorithms to classify or predict the incoming traffic. If any malicious behaviour is found then that particular 'set of hosts' is forwarded to next module. Function 2 uses the service three-way handshake to find out the exact results. It checks for open connections to state which host is an intruder.

The service three-way handshake works as: host requests the server to 3-way handshake, for which server replies as acknowledgement bit which is SYN flag bit, to this the host will reply as SYN-ACK bit thus, completing the three-way handshake. Then host which will not be ready to reply the SYN flag bit is an intruder.

The figure 2 and figure 3 below shows the flow diagram for both the algorithms:

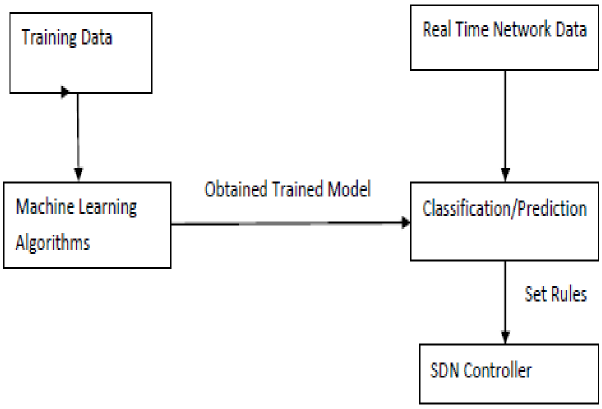

Fig. 2: Flow diagram for training the Machine learning algorithms to define set rules for controller.

5. Conclusion

SDN is emerging paradigms that have various advantages like manageable, cost effective and easy control and adaptability. DDoS attack poses a major threat to the security in controller of SDN, as controller is centralized; threat to the controller poses a threat to the entire system and its resources. The paper presents the thorough literature review of SDN, security issues and effective machine learning algorithms.

In this paper, we try to present proposed methodology for IDS that better predicts the anomalous nodes and provide better detection rate with more accuracy than the already designed algorithms.

Further, the implementation of proposed methodology using machine learning algorithm and comparisons that shows better result will be used to as the advanced IDS, as providing security to the controller which it the main goal of our work thus, increasing the processing speed SDN.

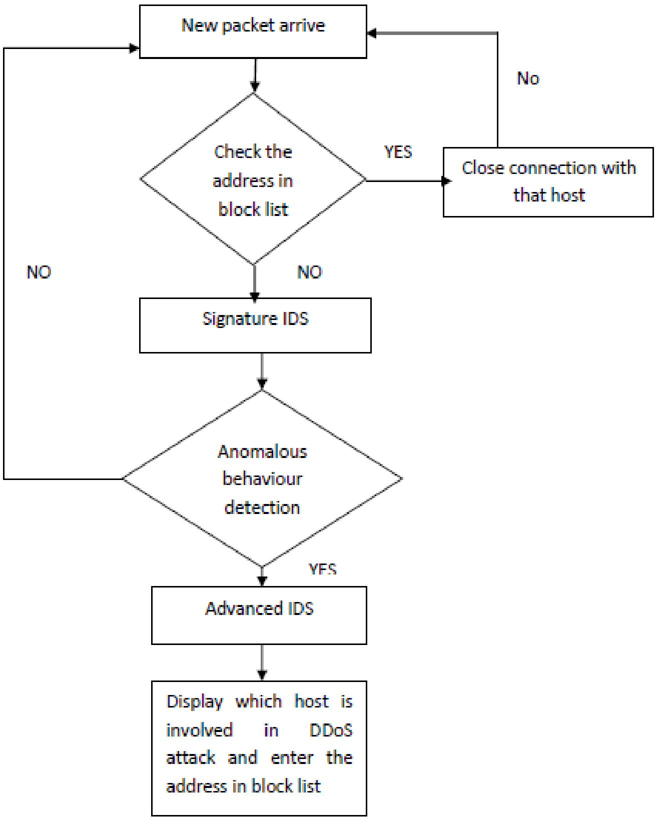
New packet arrive
No
Check the address in block list
YES
Close connection with that host
NO
NO
Signature IDS
Anomalous behaviour detection
YES
Advanced IDS
Display which host is involved in DDoS attack and enter the address in block list

Fig. 3: Flow Diagram for the implementation IDS to detection of DDoS Attack.

References

[1] Peng Zhang, Huanzhao Wang, Chengchen Hu, and Chuang Lin, “On Denial of Service Attacks in Software Defined Networks”, Network Forensics and Surveillance for Emerging Networks IEEE Network, Nov-Dec, 2016.

[2] Moreno Ambrosin, Mauro Conti, Fabio De, Nishanth Devarajan, “Amplified Distributed Denial of Service Attack in Software Defined Networking”, IEEE 2016.

[3] Lohit Barki, Amrit Shidling, Nisharani Meti, Narayan D G and Mohammed Moin Mulla, “Detection of Distributed Denial of Service Attacks in Software Defined Networks”, 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India, Sept. 21-24, 2016.

[4] Raphael Durner, Claas Lorenz, Michael Wiedemann, Wolfgang Kellere, “Detecting and Mitigating Denial of Service Attacks against the Data Plane in Software Defined Networks”, IEEE 2017.

[5] Damian Jankowski, and Marek Amanowicz, “On Efficiency of Selected Machine Learning Algorithms for Intrusion Detection in Software Defined Networks”, International Journal of Electronics and Telecommunications, 2016.

[6] Saurav Nanda, Faheem Zafari, Casimer DeCusatis, Eric Wedaa and Baijian Yan, “Predicting Network Attack Patterns in SDN using Machine Learning Approach”, IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN), 2016.

[7] S. Umarani, D. Sharmila, “Predicting Application Layer DDoS Attacks Using Machine Learning Algorithms”, World Academy of Science, Engineering and Technology, IJCEACIE, Vol. 8, No.10, 2014.

[8] Mohammad Reza Parsaei, Mohammad Javad Sobouti, Seyed Raouf khayami, and Reza Javidan, “Network Traffic Classification using Machine Learning Techniques over Software Defined Networks” IJACSA, Vol. 8, No.7, 2017.

[9] LongTail, “LongTail Log Analysis Dashboard”. http://longtail.it.marist. edu/honey/dashboard.shtml. [Online; accessed 22-April-2016].

[10] T. Padmapriya, V.Saminadan, “Performance Improvement in long term Evolution-advanced network using multiple imput multiple output technique”, Journal of Advanced Research in Dynamical and Control Systems, Vol. 9, Sp-6, pp: 990-1010, 2017.

[11] S.V.Manikanthan and V.Rama“Optimal Performance Of Key Predistribution Protocol In Wireless Sensor Networks” International Innovative Research Journal of Engineering and Technology ,ISSN NO: 2456-1983,Vol-2,Issue –Special –March 2017.

[12] Harikishore Kakarla, Madhavi Latha M and Habibulla Khan, “Transition Optimization in Fault Free Memory Application Using Bus-Align Mode”, European Journal of Scientific Research, Vol.112, No.2, pp.237-245, ISSN: 1450-216x135 /1450-202x, October 2013.