DEFENSE MECHANISM FOR DDoS ATTACK THROUGH MACHINE LEARNING

Sujay Apale[1], Rupesh Kamble[2], Manoj Ghodekar[3], Hitesh Nemade[4], Rina Waghmode[5]

[1]*Student, Department of Computer Engineering, AISSMS COE, Pune, India*
[2]*Student, Department of Computer Engineering, AISSMS COE, Pune, India*
[3]*Student, Department of Computer Engineering, AISSMS COE, Pune, India*
[4]*Student, Department of Computer Engineering, AISSMS COE, Pune, India*
[5]*Professor, Department of Computer Engineering, AISSMS COE, Pune, India*

Abstract

There is a huge advancement in Computer networking in the past decade. But with the advancement, the threats to the computer networks are also increased. Today one of the biggest threats to the computer networks is the Distributed Denial of Service (DDoS) flooding attack. This paper emphasizes the application layer DDoS flooding attacks because these (layer seven) attacks are growing rapidly and becoming more severe problem. Many researchers used machine-learning techniques for intrusion detection, but some shows poor detection and some methods take more training time. From a survey, it is found that Naïve Bayes (NB) algorithm provides faster learning/training speed than other machine learning algorithms. Also it has more accuracy in classification and detection of attack. So we are proposing a network intrusion detection system (IDS) which uses a machine learning approach with the help of NB algorithm.

Keywords: *DDoS (Distributed Denial of Service) flooding attack, Machine Learning, Naïve Bayes, Network Intrusion Detection*

----------***----------

1. INTRODUCTION

The huge advancement and rapid growth in the internet and networking has taken this computer era to a whole new level. However, this highly connected computer era has a soft spot: The hackers and attackers intentionally or non-intentionally take down some server system. Either way it financially costs too much to the company or organization whose server is under attack. To avoid damage and its cost a tool called as Intrusion Detection system is used as a last line of defense against intruders who can have unauthorized access to the system. Intrusion detection system gives the assurance of service continuity and data security. The intruder which firewall fails to detect is detected by the IDS. Even if firewall and IDS are related to network security, an IDS varies from a firewall in that a firewall looks for intrusions outwardly to stop those attacks from affecting the system by limiting the access between networks and do not give warning signal about an attack from inside. An IDS assesses a doubted intrusion which has taken place and raises an alarm. An IDS also keeps an eye on attacks that originate from inside of system.

[4]Since the summer of 1999 several DDoS flooding attacks had been launched on different organizations' web servers. The first major DDoS flooding attack occurred, in February 2000, on YAHOO in which all the services provided by company went offline for about two hours which caused an immense loss in advertising revenue of company. In October 2002, Domain Name service went offline for about an hour due to DDoS flooding attack. In February 2004, the website of SCO Group was attacked. On September 18, 2010, in USA a website of MPAA was inaccessible to internet users for over twenty hours of time because of DDoS flooding attack.

The remaining paper is structured as: Section 2 describes the classification of IDS. Section 3 categorizes the different types of application layer DDoS flooding attack. In section 4, some papers in literature are surveyed. Section 5 introduces to Naïve Bayes algorithm. Section 6 proposes an efficient intrusion detection system based on machine learning technique. Section 7 concludes the paper.

2. CLASSIFICATION OF INTRUSION DETECTION SYSTEM

IDS are of two types: Host based and Network based.

1. Host Intrusion Detection System (HIDS): HIDS run on network devices or different hosts. A Host Intrusion Detection System keeps tabs on the inward bound and outward bound packets from the device and will alert the admin if doubtful activity is spotted. It takes a snap of existing system files and compares it to the previous one. If the critically important system files were altered or deleted, the admin is alerted for investigation.
2. Network Intrusion Detection system (NIDS): NIDS are deployed at strategic points within the network to keep tabs on traffic coming in and going out from all network devices. It analyses traffic on the whole subnet and matches it with the traffic passed on the

subnets to the library of known attacks. The administrator is immediately alerted when the attack is detected.

All IDS use any one of the following techniques for intrusion detection:

1. Anomaly based IDS: These types of IDS will keep an eye on network activity and compare it with recognized baseline-data. The baseline-data will identify normal traffic for that network. Also it will identify normally used bandwidth, protocols and ports and alert the admin when traffic is detected which is atypical, or considerably different as compared to baseline-data. The problem is that it raises a false positive alarm for a genuine user if the baseline-data is not configured intelligently.
2. Signature based IDS: A signature based IDS will keep an eye on packets in the network and compare them with a signatures database or features of previously known threats. But the similarity between them is that most of the antivirus software detects malware. The problem is that there will be a delay between a novel threat being discovered and the signature for identifying that threat being applied to IDS. During that delay gap IDS can't detect new threat.

3. CATEGORIZATION OF APPLICATION LAYER DDoS FLOODING ATTACK

The application layer attacks cause exhaustion of server resources and thus cause the disruption in legitimate user's services. Application-level DDoS attacks use low bandwidth. These attacks look stealthier in appearance as they are very similar to benignant network traffic. They are non-volumetric. The most common attacks at application layer are DNS amplification flooding attack and SIP flooding attack. While major types of recent DDoS flooding attacks are those which use HTTP protocol.

1. Reflection based flooding attacks: In these attacks, attacker sends forged application layer protocol requests to large number of reflectors. Two main attacks in this category are SIP flooding and DNS amplification attacks.
2. HTTP flood attacks: It consists of seemingly legitimate session-based sets of HTTP_GET or _POST requests sent to a victim web server. These requests are consume a major amount of the server's resources. It can result in DoS without essentially needing a high-rate of traffic in the network. These types of requests are every so often sent all together by means of a number of bots, increasing the intensity of the attack.

4. LITERATURE SURVEY

In paper [1] authors proposed a neural network approach. A MLP is used for detection of intrusion, established on an off-line analysis method. This research targets to resolve a multi-class problem in which the different attack type is also identified by the neural network besides detecting whether it is a normal request or an attack. To find the optimal neural network, various neural networks are surveyed, with respects to the number of unseen layers. An early ending validation is also applied in the learning/training stage to gain the increase in the capability of the neural network generalization. The results describe that the given system classifies the records with about 91% accurateness with two unseen neuron layers and 87% accurateness with one unseen layer in the neural network.

Paper [2] suggests a layered framework combined with neural network to build an effective intrusion detection system. This system has been tested with Knowledge Discovery & Data Mining (KDD) 1999 dataset. The comparison of the systems is done with the current techniques which either use neural network layered framework. The outcome indicates that the proposed system has high attack detection accuracy and less false alarm rate. The results show that there is still opportunity to improve results as the given systems are not able to detect each attack, so it is encouraging to consider investigating in this path.

In paper [3] authors applied two of the efficient data mining algorithms called Naive Bayes and trees augmented Naive Bayes for detecting the intruders in the network and the results are compared with decision tree and SVM. They presented experimental results on NSL-KDD data set and then observed that their intrusion detection system has higher detection rate and lower false positive rate. According to the results, Naive-Bayes is found less time consuming. TAN has better accuracy rate and detection rate, and also has less false positive rate.

The paper [4] classifies the different DDoS attacks based on the deployment location, time at which they are detected, etc. Depending on these types different IDS types are categorized. This paper proposes a hybrid IDS, which is cannot be applied practically now but may be in future. But this paper also tells that the application layer DDoS flooding attacks is the largest threat because they are increasing speedily. They are stealthier as compared to DDoS attacks at other layers and they masquerade as flash crowds.

The authors in paper [5] discuss the variations in network-based and host-based intrusion detection approaches to show the together can provide additionally effective detection and prevention of intrusion. They propose a hybrid IDS combining host IDS and network IDS, with misuse detection anomaly detection techniques, uses few auditing programs to sort an wide-ranging feature set that describes host session or every network connectivity, and applies data mining to study guidelines that precisely capture the behavior of intruders and normal users. But there are still many practical and theoretical problems to be fixed, and many significant technologies are needed to study deeper. The experimental research shows that the design and implementation of accurate & efficient IDS built on data-mining is big and difficult project.

Authors employed an FC-ANN method in [6] to solve weaker detection stability and the lower detection accuracy issues with the use of restore point. In this paper fuzzy clustering technique is used to classify dataset into several subsets. These different subsets are used to train dataset. Then ANN learns the pattern of every subset. ANN is feed forward network consists of neuron with each neuron having independent processing unit. To reduce the complexness and subset size, different training subsets are generated by fuzzy clustering. Different ANN models are trained using those subsets and at last results are merged.

In paper [7], Devikrishna K. S. and Ramakrishna B. B. proposed a system using Multi-Layer Perceptron (MLP). Artificial neural network consists of neurons. Each neuron is an autonomous processing unit. The output from every neuron is sent to the neuron of next layer. In neural network input parameters consist of information extracted from network connection and output parameter class of connections like normal or attack. In this paper Multilayer Perceptron is used for intrusion detection. In this system input is mapped to appropriate output. After detecting the attack, attack is classified in to 6 types by different layers of neuron. Authors pointed out the problem of obtaining irrelevant result and suggest solving it in future work.

Numerous concerns came up from this study such as large training time, incorrect detection, more false positive rate, attack classification etc. It is essential to use high-speed machine learning technique for IDS, to solve the problem of training time and comparing the results with existing machine learning techniques. In this survey, a technique is proposed which will lessen the training time and accurateness of detection.

5. NAÏVE BAYES (NB) CLASSIFIER ALGORITHM

Naïve Bayes (NB) is the probabilistic classifier. It is based on the Bayes' theorem, in probability theory and statistics, with strong independence assumptions between different features related to a particular dataset. Simply it assumes that the existence of a particular property of a class is unrelated to the existence of any other property. It outperforms other classification techniques such as random forest, boosted trees, decision tree, etc. Methods such as clustering and nearest neighbor are mostly used with numeric data. However, data related to networks use categorical values like protocol_type, service, logged_in, etc. Advantage of using Naïve Bayes is that it requires a small database for training purpose. It is not sensitive to irrelevant features.

Bayes Theorem which is used by Bayesian Classifier states:

$$P(sj|r) = p(r|sj)p(sj)/p(r)$$

- $p(sj|r)$ = probability of instance r being in class sj

This needs to be computed.

- $p(r|sj)$ = probability of causing r in sj

We can imagine as- r is in sj, causes us to feature r with some probability.

- $p(sj)$ = probability of sj's occurence

This is just how frequent the class sj, is in the given dataset.

- $p(r)$ = probability of r's occurrence

This can actually be ignored, since it is same for all classes.

6. PROPOSED SYSTEM

It is observed that, from the survey of papers in the literature, there are some issues such as time-consuming training, low detection, less accuracy in the detection and classification of attacks, etc. So, we must find some other approach which can work on these problems. In theory, it is found that Naïve Bayes (NB) algorithm provides fast learning/training speed than existing machine learning algorithm. Therefore the proposed approach is to build an analytical model for intrusion detection which will have a fast learning/training ability than any other existing approach. Using NB method a classifier will be built to differentiate between usual and unusual activity. The results of NB algorithm will be compared with existing intrusion detection approach.

The proposed architecture for the IDS:

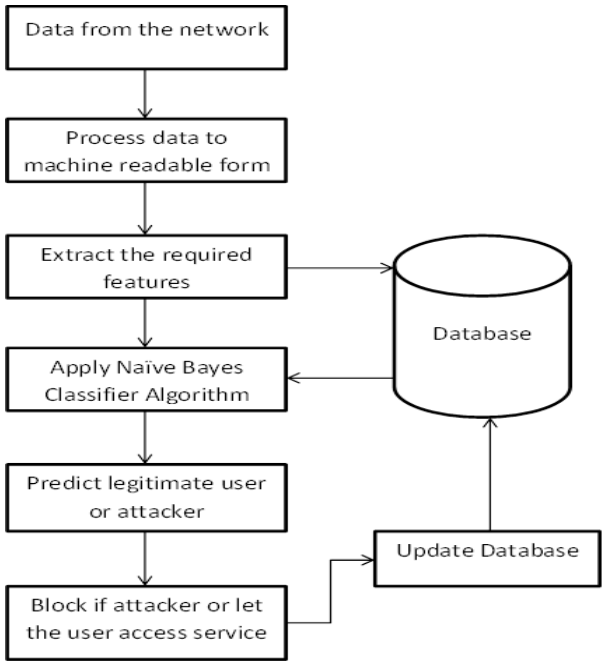

Fig 1: Proposed Machine Learning Approach for Intrusion Detection

7. CONCLUSION

In this paper we have proposed the architecture for network intrusion detection using machine learning approach. The paper mainly focuses on the application layer DDoS flooding attack. Categorization of application layer DDoS attack is given in this paper. Also we discussed different types of IDS. Various problems in the performance of the existing approaches of intrusion detection are pointed out. And to overcome these problems we propose the use of Naïve Bayes classifier algorithm for machine learning as it can improve time required to train IDS. The results of this system will be compared, with existing approaches, in the future.

ACKNOWLEDGEMENTS

This paper involves number of respected helping hands. We are grateful to Prof. Rina Waghmode for her valuable guidance. We would like to thank the Department of Computer Engineering, AISSMS COE, Pune for their uninterrupted help and support.

REFERENCES

[1] M. Moradi, M. Zulkernine, "A Neural Network Based System for Intrusion Detection and Classification of Attacks"

[2] Nidhi Srivastav, Rama Krishna Challa , "Novel Intrusion Detection System integrating Layered Framework with Neural Network", IEEE, 2012

[3] R. Najafi,Mohsen Afsharchi,"Network Intrusion Detection Using Tree Augmented Naive-Bayes", IEEE Iran Section, 2012

[4] Saman Taghavi Zargar, James Joshi and David Tipper, " A Survey of Defense Mechanisms Against Distributed Denial of Service (DDoS) Flooding Attacks", IEEE Communications Surveys & Tutorials, Ieee, 2013

[5] Duanyang Zhao, Qingxiang Xu, Zhilin Feng, "Analysis and Design for Intrusion Detection System Based on Data Mining", 2010 Second International Workshop on Education Technology and Computer Science, IEEE, 2010

[6] Prof. D.P. Gaikwad, Sonali Jagtap, Kunal Thakare, Vaishali Budhawant, "Anomaly Based Intrusion Detection System Using Artificial Neural Network and fuzzy clustering", International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 1 Issue 9, November- 2012

[7] Devikrishna K. S., Ramakrishna B. B., "An Artificial Neural Network based Intrusion Detection System and Classification of Attacks", International Journal of Engineering Research and Applications (IJERA), ISSN: 2248-9622,Vol. 3, Issue 4, Jul-Aug 2013, pp. 1959-1964

[8] V. JaiGanesh, Dr. P. Sumathi, "An Efficient Intrusion Detection using Fast Hierarchical Relevance Vector Machine", Journal of Theoretical and Applied Information Technology (JATIT), ISSN: 1992-8645, 10th April 2014. Vol. 62 No.1

[9] V. Jaiganesh, S. Mangayarkarasi, Dr. P. Sumathi, "Intrusion Detection Systems: A Survey and Analysis of Classification Techniques", International Journal of Advanced Research in Computer and Communication Engineering, ISSN (Print): 2319-5940, ISSN (Online): 2278-1021, Vol. 2, Issue 4, April 2013

[10] Kok-Chin Khor, Choo-Yee Ting and Somnuk-Phon Amnuaisuk, "From Feature Selection to Building of Bayesian Classifiers: A Network Intrusion Detection Perspective", American Journal of Applied Sciences 6 (11): 1948-1959, 2009 ISSN 1546-9239 © 2009 Science Publications

[11] InfosecInstitutes: http://resources.infosecinstitute.com/layer-seven-ddos-attacks/

[12] DDoSAttackProtection: http://ddosattackprotection.org/blog/layer-7-ddos-attack/

BIOGRAPHIES

Sujay Apale is a student at AISSMS COE, Pune. He is pursuing Bachelor's Degree in Computer Engineering in Savitribai Phule Pune University, Pune, Maharashtra, India.

Rupesh Kamble is a student at AISSMS COE, Pune. He is pursuing Bachelor's Degree in Computer Engineering in Savitribai Phule Pune University, Pune, Maharashtra, India.

Manoj Ghodekar is a student at AISSMS COE, Pune. He is pursuing Bachelor's Degree in Computer Engineering in Savitribai Phule Pune University, Pune, Maharashtra, India.

Hitesh Nemade is a student at AISSMS COE, Pune. He is pursuing Bachelor's Degree in Computer Engineering in Savitribai Phule Pune University, Pune, Maharashtra, India.

Rina Waghmode received the BE degree in IT in 2009 and ME degree in IT in 2013 for her work in Software Cost Estimation, from Pune University. She is professor of Computer Engineering at AISSMS COE, Pune. She has published 6 papers. Latest paper is published in 4th IEEE IACC 2014, Gurgaon-Delhi.