ISSN:0975-9646

Denial-Of-Service Attack Detection Technique Derived From Statistical Traffic Analysis

Mrs. Sanjivani Sumant, Prof. Sweta Kale

Department of Information Technology
R. M. D. Sinhgad School of Engineering
Sawitribai Phule Pune University, India

***Abstract*—** Denial of Service (DoS) attacks are one type of aggressive and threatening type of behaviour specially to online servers. These different types of DoS attacks severely degrade the availability of a victim, which can be a host, a server, a router, or an entire network. They enforce exhaustive computation tasks to the victim by exploiting its system vulnerability or flooding it with huge amount of useless packets. The victim can be forced out of service from a few minutes to even several days. This causes serious damages to the services running on the victim. Therefore, effective detection of DoS attacks is essential to the protection of online servers. Work on DoS attack detection mainly focuses on the development of network-based detection mechanisms. Detection systems based on these mechanisms monitor traffic transmitting over the protected networks. The proposed system focuses on the NSL KDD dataset traffic records and extract the features which are directly associated with DoS attacks. Based on these features, the statistical traffic analysis like Triangular Area Map (TAM) generation and Multivariate Correlation Analysis (MCA) are computed for normal profiles. These models are used as reference models to detect any known as well as unknown DoS attacks in the network with enhanced Detection rate and True positive rate (TPR).

Keywords— **DoS, victim, Statistical Traffic Analysis, TAM, MCA, Detection Rate, TPR**

I. INTRODUCTION

Denial of Service (DoS) attacks are one type of aggressive and threatening type of behaviour specially to online servers. These different types of DoS attacks severely degrade the availability of a victim, which can be a host, a server, a router, or an entire network. They enforce exhaustive computation tasks to the victim by exploiting its system vulnerability or flooding it with huge amount of useless packets. The victim can be forced out of service from a few minutes to even several days. This causes serious damages to the services running on the victim.

There are multiple techniques which deal with Denial-of-Service attacks which will detect and can also mitigate risk caused by the DoS attacks. There are two main types of DoS attack detection techniques, Misuse-based detection technique and Anomaly-based detection technique. In Misuse-based detection technique, attack is detected by monitoring the network activities and matching it with the existing attack signatures. This can better detect the known attacks and there is administrative overhead of uploading and updating the signatures and also requires network and security expertise[2,3]. In Anomaly-based detection technique, system monitors the network activities, prepares the normal profile threshold based on legitimate traffic. This can be done by using different machine learning or statistical analysis [7,8]. The deviation of the observed attack traffic can be further compared with the threshold or normal profile traffic; if the deviation is within the acceptable range then it can be declared as normal traffic or can be declared as an attack. With this approach, it can be possible to detect zero day attack.

The main motivation behind this work is that recently many organizations in the world experienced the serious denial of service attacks. Some of them are [10]:

- JPMorgan Chase and Citigroup suffered intermittent online disruptions, according to Fox Business. According to tweets posted in August 2013, Chase and Citi both acknowledged suffering site issues.
- The Chase banking website appears to have been unavailable from 8:55 a.m. ET until 10:21 a.m. ET," he says. "Our monitoring agents reported DNS [Domain Naming System] lookup errors throughout that period, across the U.S."
- DNS is the system that translates a website's name, such as www.chase.com, into an Internet protocol address that's assigned to a Web server for that site, Rudger explains.
- The latest attacks have been erratic and seemingly less targeted. U.S. bank ing institutions, which have been under attack since September 2012, have adapted their defences, making their online-banking sites hard to take down, experts say [10].

So to detect the attacks on targets, which aims at exhausting target resources, thereby denying service to valid users is the main aim.

In this system, the Anomaly-based denial of service attack detection system is designed which is based on statistical analysis. The statistical analysis techniques such as Multivariate Correlation Analysis and Triangular Area Map generation are used to find the geometrical correlation between the different attributes of the packet. These different attributes of the packets are normalized and then the MCA and TAM are calculated. The Euclidian Distance ED is calculated for all the TAMs. The Normal Profile is generated and threshold range is calculated for legitimate traffic. Attack traffic is verified against the threshold range for detection of DoS attack. Again same procedure is carried out for attack traffic i.e. attributes normalization, generation of MCA and TAM etc.

The evaluation of the system is carried out by using the KDD Cup 99 and NSL-KDD dataset. In these datasets the DoS attack related features are captured and evaluation is carried out.

II. RELATED WORK

A lot of work had been done on Denial of Service attacks. The main objective is to promptly detect the attack i.e to get more detection rate and to minimize the False Negative Rate.

A. *Misuse Type detection system.*

In this paper, Misuse Type detection system is addressed which works on BRO which is conceptually divided into "BRO Event Engine" and "Policy Script Interpreter". BRO Event Engine reduces a stream of _altered packets into stream of higher-level network events and Policy Script Interpreter interprets the events to express site's security policy and generate a syslog. This analyses four applications: Finger, FTP, Portmapper and Telnet [12].

B. *Anomaly Type Intrusion Detection System.*

This is a Real Time Intrusion Detection System: Detected by monitoring system's Audit records for abnormal patterns. It covers not only DoS attack but other intrusions also like Virus, Trojan Horse. This is a general Purpose Intrusion Detection System which is independent of particular system, Application or type of intrusion. Statistical based, Knowledge based and Machine Learning based techniques are discussed in this paper. This describes the main features of several currently available IDS systems [5].

C. *Different Anomaly based detection technique like statistical, knowledge based and Machine Learning.*

This paper gives the survey of different Anomaly based detection techniques like Statistical based, Knowledge based and Machine Learning based techniques. Also discussed pros and cons of these techniques.

1. Statistical Based Technique - Prior knowledge of activities is not required but it is difficult to set the parameters and Matrices.
2. Knowledge Based Technique - These techniques are flexible and scalable but it is difficult to prepare knowledge base from Data.
3. Machine Learning Based Technique - These techniques are also flexible and scalable but it is highly dependent on assumptions and also resource consuming [5].

D. *Multivariate Correlation Analysis (MCA) based nonpayload based DoS detection approach which uses Triangular Area Map (TAM) generation technique for individual record*

The Anomaly Based Intrusion Detection system in which MCA is used along with triangular area map generation. A TAM is constructed and triangle areas are arranged on maps. Two TAM s are compared for investigation. Mahalanobis Distance which is used mainly to measure the dissimilarities between traffic records [2,3]

This MD is used for Normal profile generation.(Algorithm). Threshold value is calculated and attack is detected. This particular system is implemented in the proposed system for detection of DoS attack [1, 2]. But it needs improvement in its Detection Rate and False Positive Rate.

E. *For evaluating the performance of DoS detection technique, KDD Cup 99 Dataset is used which can cover seven types of DoS attacks.*

All the evaluation is done by using 10 percent of KDD Cup99 Dataset but the main aim is to test this system using Real time data with improved false positive rate of detection. Anomaly Based Detection- KDD Cup 99 Data set is widely used for Anomaly Based detection. Derived from DARPA'98 Dataset.KDD Cup 99 dataset can address Seven types of DoS attacks i.e. Normal, Teardrop, Neptune, Smurf, Pod, Neptune, Land and Back [4].

In this system, Anomaly based Intrusion detection technique is designed which uses statistical approaches for attack detection. Also the work had been carried out on NSL-KDD dataset. The scope is limited to TCP packets only.

III. PROPOSED WORK MODEL

The proposed work for Denial of Service attack detection is carried out using NSL KDD dataset. The dataset consist of the following 24 statistical features; 14 conventional features and 10 additional features. The first 14 features were extracted based on NSL KDD data set, which is a very popular and widely used performance evaluation data in intrusion detection research field. With this approach sample by sample detection is possible. Each sample can be investigated separately and attack can be detected.

There are basic three steps involved in the proposed system Architecture.

A. *Step 1: Basic Feature selection for Individual Record*

The NSL-KDD data set suggested to solve some of the inherent problems of the KDDCUP'99 data set [11].

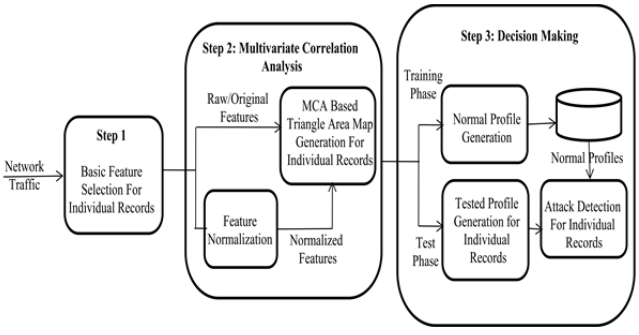

Fig 1. Proposed system Architecture

KDDCUP'99 is the mostly widely used data set for anomaly detection. But Tavallaee et. al. conducted a statistical analysis on this data set and found two important issues that greatly affected the performance of evaluated systems, and results in a very poor evaluation of anomaly detection approaches. To solve these issues, they proposed a new data set, NSL-KDD, which consists of selected records of the complete KDD data set .

There are some advantages of NSL KDD dataset over KDD CUP99 dataset. it does not include redundant records in the train set, so the classifiers will not be biased towards more frequent records. Second, the number of selected records from each difficulty level group is inversely proportional to the percentage of records in the original KDD data set [11].

From the 41 features from NSL KDD dataset 12 conventional features are selected for detection of DoS attack. The main features like Duration, Service, Source Bytes, Destination Bytes, Count, Same Srv Rate, Serror Rate, Srv serror Rate, Dst host count, Dst host srv count, Dst host same src port rate, Dst host serror rate, Dst host srv serror rate are taken for evaluation purpose.

These different packet features are given as an input to the second step, the Multivariate Correlation Analysis step.

B. Step2: Multivariate Correlation Analysis

This step again consists of two steps, Feature Normalization and Multivariate Correlation Analysis based Triangular Area Generation.

1) Feature Normalization

All the features selected in first step can vary in its attribute values. So to achieve better detection rate and desired results, features need to be 'Normally' distributed [2,3].

For this purpose the statistical distribution i.e. Normal Distribution is used. With Normal Distribution 'Bell Curve' is achieved and Mean and Standard Deviation is calculated for each and every attribute.

The Probability density of the normal distribution is

$$f(x,\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

We can get the normalized attributes as a output of this step

2) MCA based Triangle Area Map Generation

DoS attack traffic behaves differently from the legitimate network traffic, and the behaviour of network traffic is reflected by its statistical properties. To well describe these statistical properties, we present a novel MCA approach in this section. This MCA approach employs triangle area for extracting the correlative information between the features within an observed data object (i.e., a traffic record) [2,3].

For each and every sample of the network traffic, Triangular Area is calculated between every combination of attributes. If x_i is a packet containing features or attributes $x_i = (f_1^i, f_2^i, f_3^i, ...f_k^i)$, Triangular Area Map $Tr_{j,k}^i$ is calculated between every two features $\Delta f_i^j O f_i^k$ by following equation [6]

$$Tr^i_{j,k} = | f_i^j | x | f_i^k| /2$$

When all the TAMs are calculated for one packet, a symmetrical matrix is generated. The lower triangle of this matrix is taken or evaluation as this is a symmetric matrix.

The output of this module is an array of lower matrix TAMs.

$$TAM^i_{lower} = [Tr^i_{2,1}, Tr^i_{3,1}, Tr^i_{2,3}.....Tr^i_{m,m-1}]$$

Thus for each and every normal or legitimate packet TAM^i_{lower} is calculated and mean of TAM^i_{lower} is calculated which is further used for calculating 'Euclidian Distance'.

C. STEP 3: DECISION MAKING

1) Normal Profile Generation

For legitimate packets in dataset, the 'Euclidian Distance' is calculated between every packet's TAM^i_{lower} and the mean $\overline{TAM^i_{lower}}$. Once 'Euclidian Distance' is calculated for every normal packet [9].

$$ED(\overline{TAM^i_{lower}}, TAM^i_{lower}) = \sqrt{(TAM^i_{lower} - \overline{TAM^i_{lower}})^2}$$

and mean of all 'Euclidian Distance', 'μ' is calculated. Also standard deviation 'σ' is calculated .

As Normal Distribution is used for feature normalization, value 'α' ranges from 1 to 3. These 'μ', 'σ' and 'α' are used to define the threshold value. Threshold value ranges from 'μ + σ *α' and 'μ - σ *α' This range is the Normal Profile range.

2) Test Profile Generation

For all the attack packets in dataset, the features are normalized, TAMs are calculated and 'Euclidian Distance' is calculated for each and every sample of packet.

This 'Euclidian Distance' is validated against the Normal Profile Threshold range. If this is within the range of Threshold then it can be detected as 'Normal Traffic packet' or it will be detected as an 'Attack'

Requirement - $TAM^{observed}_{lower}$, Normal Profile TAM^{normal}_{lower}

1. Generate $TAM^{observed}_{lower}$ for the observed traffic record.
2. Calculate

   $ED^{observed} = ED\,(TAM^{observed}_{lower}, \overline{TAM^{normal}_{lower}})$

3. If $(\mu - \sigma * \alpha) \leq ED^{observed} \leq (\mu + \sigma * \alpha)$
4. Then return **Normal.**
5. Else
6. Return **Attack**
7. End if

Fig 2. Algorithm for attack detection based on Euclidian distance.

RESULT ANALYSIS OF NORMALIZED DATA

Evaluation of attack detection is done by using NSL KDD dataset. Normal Profile is built by using NSL KDD Training dataset. Test profile is generated by using NSL KDD Test dataset. The Euclidian Distance is calculated for both Normal and Test Profiles. Threshold range is generated by using 'μ + σ *α' and 'μ - σ *α' For normal Distribution, the value of 'α' ranges from 1 to 3. Detection rate and False positive rate is evaluated for the different values of 'α'.

The Detection rate of DoS attack by using NSL KDD dataset is shown in following table.

	Threshold				
Type of Record	1α	1.5α	2α	2.5α	3α
Normal	100 %	100 %	100 %	100 %	100 %
Teardrop	100 %	100 %	100 %	100 %	100 %
Smurf	100 %	100 %	100 %	100 %	0 %
Neptune	100 %	100 %	100 %	100 %	100 %
Land	100 %	100 %	100 %	100 %	100 %
Back	100 %	100 %	100 %	98.04%	97.06

Table 1. Detection Rate of NSL KDD Test Data

The Overall DR and FPR for NSL KDD text dataset is as follows.

	Threshold				
Type of Record	1α	1.5α	2α	2.5α	3α
DR	100%	100%	100%	99.61%	79.41%
FPR	0%	0%	0%	0%	0%

Table 2. Detection Rate and False Positive Rate of NSL KDD Test Data

The graphical representation of Detection Rate for different values of 'α'

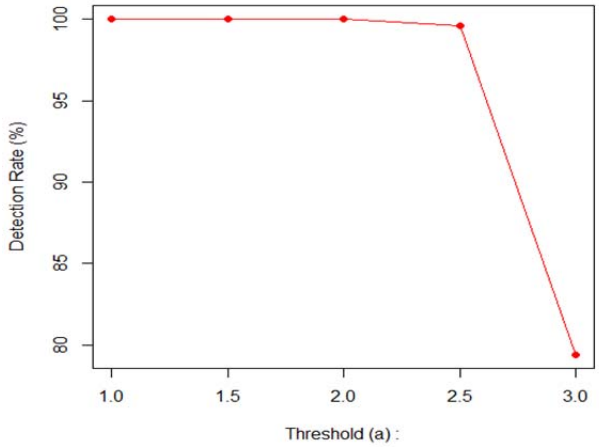

Fig. 3. Curve for detection DoS attack.

Thus it is possible to achieve almost 100 % Detection Rate and 0 % False Positive Rate.

CONCLUSION AND FUTURE WORK

With this paper a DoS attack detection system, which is a TAM based MCA technique, it is possible to evaluate each and every packet to detect the attack. So sample by sample evaluation of the data is possible. With the different values of 'α' between 1 to 3 for Normal Distribution of data, we can achieve almost 100 % Detection Rate and zero False Positive Rate.

As a part of future work, we can verify the Detection Rate and False Positive Rate for various 'σ' values. Also we can further test the system with Real Time Data by injecting the attack by using DoS attacking Tools.

ACKNOWLEDGMENT

I would like to thank my Guide and H.O.D. Prof. Sweta Kale for her precious contribution and extended support throughout the project work .

REFERENCES

[1] Sanjivani Sumant, Sweta Kale, "Overview of Denial-Of-Service Attack and statistical detection Techniques", IJERT Volume. 3 , Issue. 11 , November - 2014.

[2] Zhiyuan Tan; Jamdagni, A.; Xiangjian He; Nanda, P.; Ren Ping Liu, "A System for Denial-of-Service Attack Detection Based on Multivariate Correlation Analysis," Parallel and Distributed Systems, IEEE Transactions on , vol.25, no.2, pp.447,456, Feb. 2014.

[3] Z. Tan, A. Jamdagni, X. He, P. Nanda, and R.P. Liu, "Triangle-Area-Based Multivariate Correlation Analysis for Effective Denial-of-Service Attack Detection", Proc. IEEE 11th Intl Conf. Trust, Security and Privacy in Computing and Comm., pp. 33-40, 2012.

[4] M. Tavallaee, E. Bagheri, L. Wei, and A.A. Ghorbani, "A Detailed Analysis of the KDD Cup 99 Data Set", Proc. IEEE Second Intl Conf. Computational Intelligence for Security and Defense Applications,pp. 1-6, 2009.

[5] P. Garca-Teodoro, J. Daz-Verdejo, G. Maci-Fernndez, and E. Vzquez, "Anomaly-Based Network Intrusion Detection: Techniques, Systems and Challenges," Computers and Security, vol. 28, pp. 18-28, 2009.

[6] Author Zhiyuan Tan, "Generation Of Network Behaviour Descriptor Using MCA Based On TAM", Retrieved from http : \\www:kaspersky:com=images=Zhiyuan Tan:pdf

[7] G. Thatte, U. Mitra, and J. Heidemann, "Parametric Methods for Anomaly Detection in Aggregate Traffic," IEEE/ACM Trans. Networking, vol. 19, no. 2, pp. 512-525, Apr. 2011.

[8] S. Yu, W. Zhou, W. Jia, S. Guo, Y. Xiang, and F. Tang, "Discriminating DDoS Attacks from Flash Crowds Using Flow Correlation Coefficient," IEEE Trans. Parallel and Distributed Systems, vol. 23, no. 6, pp. 1073-1080, June 2012.

[9] "Euclidian Distance", Retrieved from http://en.wikipedia.org/wiki/Euclidean_distance

[10] "ddos-phase-4-attacks-sporadic-a-6006" Retrieved from http://www.bankinfosecurity.com/ddos-phase-4-attacks-sporadic-a-6006

[11] Feature Selection for Intrusion Detection using NSL-KDD by Hee-su Chae, Byung-oh Jo, Sang-Hyun Choi, Twae-kyung Park, ACCIS-30, Recent Advances in Computer Science

[12] V. Paxson, "Bro: A System for Detecting Network Intruders in Real-Time," Computer Networks, vol. 31, pp. 2435-2463, 1999.