Detection of DoS Attack based on MCA and TAM

Miss Kanchan D. Sherkar[1] Mr. Sandip A. Kahate[2]
[1,2]Department of Computer Engineering
[1,2]SPCOE Pune

Abstract— The huge number of Internet users and services are increases. All important services which are based on Internet needs to be maintained properly so that the users can avail them whenever they need. If the services are not available in time it will create a crisis. As the numbers of hosts in Internet are increasing, the threats to it are also increasing. Denial of Service (DoS) is the most deadly threats rising in Internet. Web services require security and stability and from these two concerns there are some methods that can disorientate DoS attack from flash crowd and trace the sources of the attack in large amount of traffic in network. But it is difficult to detect the exact sources of DoS attacks in traffic of network when crowd event is also present. After analyzing the characteristics of DoS attacks and the existing Algorithms to detect DoS attacks, In this Paper proposes a detecting and tracing algorithm for DoS attacks based on Multivariate correlation Analysis and triangle map generation. The proposed approach focuses majorly on the efficiency and scalability features with minimum overhead in terms of resources and time, removal of traffic pattern dependency, increase in detection rate between DoS and flash crowd and also trace the sources of DoS attack.

Key words: Denial-of-Service Attack, DDOS Attack Network Traffic Characterization, Multivariate Correlations, Triangle Area

I. Introduction

A Distributed Denial of Service attack is commonly characterized as an event in which a legitimate user or organization is deprived of certain services, like web, mail or net-work connectivity, that they would normally expect to have. DDoS is basically a resource overloading problem. The resource can be bandwidth, memory, CPU cycles, file descriptors, buffers etc. The attackers bombard scare resource either by flood of packets or a single logic packet which can activate a series of processes to exhaust the limited resource .In the Figure.1.1 below, simplified Distributed DoS attack scenario is illustrated. The shows that attacker uses three zombies to generate high volume of malicious traffic to flood the victim over the Internet thus rendering legitimate user unable to access the service.

Distributed denial of service (DDoS) [1] has caused a huge economic loss to the victims. Therefore, the detection of traffic anomaly is important to the security of modern networks. DDOS attack and flash crowd attack are identified and blocked by detection and prevention methods. Attack prevention aims to fix security holes, such as crash up of servers by DDOS and flash crowd attack [2]. This approach aims to improve the global security level and is the best solution to DDOS attacks in theory. Both flash crowds and denial of service attacks have the potential to have similar impact on Web servers. We demonstrate a way to distinguish between them using our security model to identify the network traffic, so that Web servers can attempt to serve normal clients and drop requests from clients involved in attacks and also to block the misbehaving users. Attack detection aims to detect DDoS attacks in the process of an attack and characterization helps to distinguish attack traffic from legitimate traffic[4]. A flash event (FE) [3] is a large surge in traffic to a particular Web site causing a dramatic increase in server load and putting severe strain on the network links leading to the server, which results in considerable increase in network traffic. A denial of service attack (DoS) [5] is an explicit attempt by attackers to prevent legitimate users of a service from using that service. Consider any attempt to undermine a Web site to be a denial of service attack. Network traffic anomaly detection can be done through the self-similar analysis of network traffic.

II. Related Theory

In paper [1] author used information distance technique to distinguish DoS from flash crowd. Both these attacks are motivated different methods to measure the similarity among flows such as Abstract distance metrics, Jeffrey distance, Sibson distance, Hellinger distance. After comparison among these four metrics, it is found that the Sibson distance is the most suitable method. By applying an algorithm to the real datasets, an accuracy around 65% and it is very efficient to improve an accuracy of the flow based discrimination strategy.

In [2] DoS is distinguished from flash crowd by using probability metrics. They proposed main contributions to distinguish DoS attacks from Flash crowds as hybrid metric and the Bhattacharyya metric. The hybrid metric can reduce the false positive rate greatly. But the limitation of this method is that it is not applied in the real network situation, and so cannot find out more recognizable characteristics of IP packets.

Paper [3] presented a packet arrival pattern for distinguishing DoS from flash event. In this paper, two methods are used, first Behavior based detection which can discriminate DoS attack traffic from traffic generated by real users and second Pearson correlation coefficient which can extract the repeatable features of the packet arrivals. The major limitation is two methods are not tested with different packet information such as packet delay and changing rate of port number so that it can test with the real scenarios in real time. So there is no confirmation of the performance from the predictability test.

In [4] discrimination of DoS from flash crowd is done with the help of flow correlation coefficient, used as a similarity metric among suspicious flows. Limitations of this method are, the detection rate of differencing DoS from flash crowd is less, tracing of the sources of the DoS attack is not given and it is very hard to identify DoS attack flows at sources since the traffic is not so aggregate using world cup dataset.

III. SYSTEM DESIGN AND PROPOSED SYSTEM

The detail detection mechanism of DoS attack is given in this section, where the system framework and the sample-by-sample detection mechanism are described.

The proposed DoS attack detection system combines parameters from KDD CUP 99 dataset such as duration, protocol type, service, srcbyte, destbytes, etc, is used to detect an attack. In this way, The novel approach is to improve the global security level and is the best solution to DOS attacks in theory. Also the novel tracing system is used to trace the sources of the DDoS attack .It will detect the subtypes of DDoS attack such as amplification attack, smurf, fraggle attack etc[6].

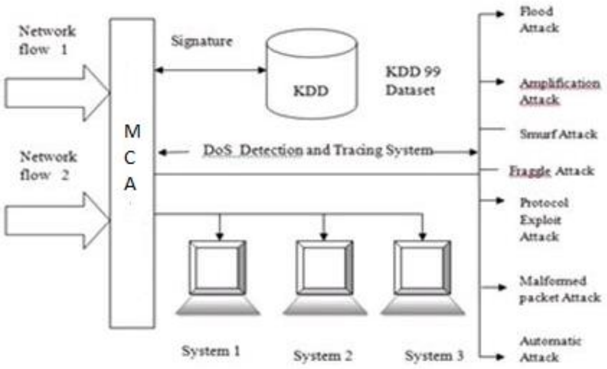

Fig. 1: System Architecture

IV. MULTIVARIATE CORRELATION ANALYSIS

Dos attack traffic different from legitimate traffic. The multiple correlations coefficient is a measure the prediction of given variable using a linear function of a set of other variables. It is measured by the square root of determination, the best possible linear predictors are used under the particular assumptions, whereas the coefficient of determination is describe for multiple general cases, which include nonlinear prediction which the predicted values have not been derived from a model-fitting procedure. The multiple correlation takes values between zero and one; a higher value indicates a better predictability of the dependent variable from the independent variables, with a value indicating that the predictions are exactly correct and a value of zero indicating that no linear combination of the independent variables is a better predictor than is the fixed mean of the dependent variable.

[7]To describe these statistical properties, present Multivariate Correlation Analysis (MCA) mechanism with triangle area for extracting the correlative information between the features within an network traffic record. [10] A Triangle Area Map(TAM) is constructed and all the triangle areas are stored on the map with respect to their indexes. *TAMi* is a symmetric matrix having elements of zero on the main diagonal. When comparing the two Triangle area map, the map into two images symmetric along their main diagonals. Any differences were identified on the upper triangles of the images, and can be found on their lower triangles as well. Therefore, to perform a quick comparison of the two TAM, to choose to investigate either the upper triangles or the lower triangles of the TAM only. This produces the same result as comparing using the entire TAM.

V. DETECTION MECHANISM

In this section, a threshold-based anomaly detector was proposed in this whose normal profiles are created using legitimate network traffic records and used for future comparisons with new incoming traffic records. The dissimilarity between a new incoming traffic record and the respective normal profile is examined by the proposed detector [8]. If the dissimilarity is greater than a pre-determined threshold, the traffic record is flagged as an attack. Otherwise, it is labeled as a legitimate traffic record. So, normal profiles and thresholds have direct influence on the performance of a threshold-based detector. [4] A low quality normal profile causes an inaccurate characterization to legitimate network traffic. Thus, we first apply the proposed triangle area- based MCA approach to analyze legitimate network traffic, and the generated TAMs are then used to supply quality features for normal profile generation.

A. Normal Profile Generation

Assume there is a set of g legitimate training traffic records Xnormal = {xnormal 1 , xnormal 2 , · · · , xnormal g }.The triangle-area-based MCA mechanism is applied to analyze the records. The generated lower triangles of the TAMs of the set of g legitimate training traffic records are denoted by Xnorm TAMlower ={TAMnormal,1lower , TAMnormal,2lower , · · · , TAMnormal,glower}. [4] Mahalanobis Distance (MD) is adopted to measure the dissimilarity between traffic records. This is because MD has been successfully and widely used in cluster analysis, classification and multivariate outlier detection techniques. MD evaluates distance between two multivariate data objects by taking the correlations between variables into account and removing the dependency on the scale of measurement during the calculation. [10]

B. Threshold Selection

The threshold is used to differentiate attack traffic from the legitimate traffic. [10] $Threshold = \mu + \sigma * \alpha$ For a normal distribution, α is usually ranged from 1 to 3. This means that detection decision can be made with a certain level of confidence varying from 68% to 99.7% in association with the selection of different values of α. Thus, if the MD between an observed traffic record xobserved and the respective normal profile is greater than the threshold, it will be considered as an attack.

C. Attack Detection

To detect DoS attacks, the lower triangle (TAMobservedlower) of the TAM of an observed record needs to be generated using the proposed triangle-area-based MCA mechanism.[10] Then, the MD between the TAMobserved lower and the TAMnormal lower stored in the respective pre-generated normal profile NormPro is computed. The detailed detection algorithm is shown in below.

VI. ALGORITHM FOR ATTACK DETECTION

1) Initialize I (no. of records) = 1
2) Scan each record of the rule pool set.
3) Find number of items (N), number of transactions (M)
4) Increment I by 1 and repeat step 2 until last record in the rule poolset

5) Initialize k (number of itemset) =1
6) Find frequent itemsetLk from Ck of all candidate itemsets Lk is data record in collected data set and Ck is data record in KDD data set Scan D and count each itemset in Ck,If count is greater than minimum support, then it is frequent
7) Form Ck+1 from Lk; k = k + 1 Join Lk-1 itemset with itself to get the new candidate itemsets, If found a non-frequent subset then remove that subset.
8) Check all rules with test dataset
9) Repeat step 6 and step 9 until Ck is empty

VII. Evaluation of DoS Attack Detection System

The experiments of this novel proposed system are performed by using KDD CUP 99 dataset. It contains four modules. At the beginning we have to capture packets from different networks and store it on any text file. Then we have to generate the rules based on KDD CUP 99 dataset.[9] After that we have to trace the DoS attack by using above explained thermos of network based on similarities. Then we have to calculate detection ratio based on correlation coefficient with the help of frequent pattern growth algorithm so that we can classify DoS

Table 1 Evaluation table attack in diffierent types. At the end we will analyse the result expecting increase in rate of detection of DoS from given

Attack name	Threshold					
	1σ		2σ		3σ	
	DT	FPR	DT	FPR	DT	FPR
Back	98.74 %	1.26 %	98.59 %	0.96 %	98.30 %	0.75 %
Land	0.00%	1.97 %	0.00%	1.45 %	0.00%	0.74 %
Neptune	82.40 %	1.77 %	55.10 %	0.92 %	51.92 %	0.69 %
Pod	99.97 %	0.57 %	99.95 %	0.54 %	99.90 %	0.30 %
Smurf	99.80 %	1.65 %	99.69 %	0.92 %	99.67 %	0.73 %
Teardrop	71.40 %	2.53 %	58.42 %	1.69 %	49.46 %	1.20 %

Table 1: Evaluation Table Event

The experiments are performed by using International Knowledge Discovery Dataset. The KDD CUP 99 dataset is publicly available and considered as a benchmark dataset for testing of various detection algorithms by using KDD CUP 99 dataset, rather than inserting the attack packets into the normal traces, the labeled attack samples which are obtained by passive monitoring [6].

The KDD CUP 99 datasets consist of two types of dataset: training dataset and testing dataset.[9] Each record of the training data is labeled as either anomalous or normal, which denotes a specific kind of attack. The training dataset contains a total 22 types of attacks and in the testing dataset, 395 dataset has contain additional 15 types of attacks [7].As we are detecting, sources of DDoS attacks (Smurf, fraggle, Neptune, Teardrop and Ping of Death). After elaborating labeled dataset, it has been found that total number of 41 attributes provides the specifications of the received packets. For this experiment, by using different attributes of packet flows such as time, duration, protocol, service, flag, source bytes, destination bytes at each router to differentiateDDoS from flash crowd.

VIII. Conclusion

This paper has presented a DoS attack detection system by using Multivariate correlation analysis mechanism with triangle map generation. The MCA mechanism extracts the geometrical relationships between individual features within each network traffic record, and offers more accurate characterization for network traffic behaviours. The evaluation result shows that the presented technique is able to detect most of the attacks and have less FPR. Further we will use machine learning techniques to enhance the detection rate.

References

[1] Shui Yu, T.Thapngam, Jianwen Liu; Su Wei, Wanlei Zhou, "Discriminating DDoS Flows from Flash Crowds Using Information Distance" Network and System Security, 2009. NSS '09. Third International Conference on 19-21 Oct. 2009

[2] Ke Li; Wanlei Zhou; Ping Li; Jing Hai; Jianwen Liu, "Distinguishing DDoS At-tacks from Flash Crowds Using Probability Metrics," Network and System Security, 2009. NSS '09. Third International Conference on 19-21 Oct. 2009.

[3] Thapngam, T.; Shui Yu; Wanlei Zhou; Beliakov, G., "Discriminating DDoS attack traffic from flash crowd through packet arrival patterns," Computer Communica-tions Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on 10-15 April 2011

[4] Shui Yu; Wanlei Zhou; WeijiaJia; Song Guo; Yong Xiang; Feilong Tang, "Discrim-inatingDDoS Attacks from Flash Crowds Using Flow Correlation Coefficient," Parallel and Distributed Systems, IEEE Transactions on June 2012

[5] Arbor, IP Flow-Based Technology, http://www.arbornetworks. com, 2011.

[6] AdaBoost-Based Algorithm for Network Intrusion Detection Weiming Hu, Senior Member, IEEE, Wei Hu, and Steve Maybank, Senior Member, IEEE.

[7] Traffic flooding attack detection with SNMP MIB using SVMq Jaehak Yu, Hansung Lee, Myung-Sup Kim, Daihee Park Department of Computer and Information Science, Korea University, Yeongi-Gun, Republic of Korea

[8] Parametric Methods for Anomaly Detection in Aggregate Traffic Gautam Thatte, Student Member, IEEE, Urbashi Mitra, Fellow, IEEE, and John Heidemann, Senior Member, IEEE.

[9] M. Tavallaee, E. Bagheri, L. Wei, and A. A. Ghorbani, “A Detailed Analysis of the KDD Cup 99 Data Set,” The The Second IEEE International Conference on Computational Intelligence for Securityand Defense Applications, 2009, pp. 1-6.

[10] A System for Denial-of-Service Attack Detection Based on Multivariate Correlation Analysis Zhiyuan Tan, Aruna Jamdagni, Xiangjian He‡, Senior Member, IEEE, Priyadarsi Nanda, Member, IEEE, and Ren Ping Liu, Me_mber, IEEE.