Denial-of-service Attack Detection System Based On Multivariate Correlation Analysis using Triangle Area Maps

Ankush Bhat , Pooja Ingole ,Rahul Ingole , Pooja Garje

Abstract— We are aware about phenomenal growth of internet .Given the unrestricted number of free websites, the internet has undeniable opened a new way of cybercrime. As one of the major network intrusive activities denials of service (DoS) attack got much attention due to its continuous growth and service impact on internet in this paper we propose a system for DoS attack detection that uses multivariate correlation analysis MCA for accurate characterization of network traffic. Although researches have been provided system based on MCA to overcome this problem there are some constrain in this work. In this paper we proposed a MCA technique based on Mahalanobis Distance (MD) the proposed system analyses original feature space(first order statistics) and extract geometrical correlation between them namely second order statistics .Proposed system use anomaly based detection to recognize attack and therefore our system is capable of detecting both known and unknown attacks effectively further triangle area map tech improves speed of MCA .Our proposed system is evaluated using KDD cup 99 data set and they have an impact on the facts on the performance of the proposed gadget examined

Index Terms—
Denial-of-Service attack, network traffic characterization, multivariate correlations, triangle area.

I. INTRODUCTION

Use of internet is increased, nowadays people use interconnected system like web servers ,cloud computing ,database servers and they need to send and receive data over the network Therefor the growing number of network intrusive activities poses a serious threat to their liability of network service .Businesses and individuals are suffering from this malicious interception . DOS attack is one of the major network intrusive active which is growing continuously and creating serious impact on the internet .DOS attack critically degrade the efficiency of a host a router or a whole network in DOS attack attacker floods the bandwidth of the victims network or fills his E-mail box with spam mail depriving him of the service he is entitled to accepted or provide. Although the means to carry out, motives for the target of a DoS attack may vary, it generally consists of the concerted efforts a person or people to prevent the internet site or service from functioning efficiently or at all temporarily or indefinitely. The goal of DoS is not to gain unauthorized access to system or data but to prevent intended users of a service from using it. A DoS attack may flood a network with traffic, thereby preventing legitimate network traffic; it may disrupt connections between two systems, thereby preventing access to a service.Therefore, effective detection of DoS attack is essential to the protection of online services.Intrusion detection systems can be classified as, Network Intrusion Detection System (NIDS) and Host Intrusion Detection System (HIDS). Host Intrusion Detection System (HIDS) is installed on workstation which are to be monitored.HIDS can monitor a workstation where it installed and unable to monitor the whole network. It is difficult to analyze the intrusion with HIDS attempted on multiple computers.Therefore work on DoS attack detection system mainly focuses on the development of Network Intrusion Detection System (NIDS). Network Intrusion Detection System (NIDS) usually consists of network appliances with a network interface card (NIC) operations in promiscuous mode and a separate management interface. The IDS is placed along a network segment and monitors all traffic on that segment.Hence, host server remains dedicated to provide quality service with minimum delay in response as it does not need to detect attacks. NIDS are future classified into(1) Signature based (knowledge based) system (2) Anomaly based (behavior based) system.Signature based intrusion detection system refers database of previous attack signature and known system vulnerabilities and hence it is not capable of detecting unknown attacks.Anomaly based intrusion detection system references a baseline or learned patterns of normal system activity to identify active intrusion attempts division from this baseline or pattern cause an alarm to be triggered. Hence it is capable of detecting both known and unknown attacks. Thus we are developing a system based on anomaly based intrusion detection system. Although this approach improves detection accuracy, it is vulnerable to attacks that linearly change all monitored features .In addition, this approach can only label an entire group of observed features but not the individuals in the group.to deal with this problem an approach based on triangle area map

Manuscript received Jan.*2016*

Ankush Bhat, *computer department, SSUP, PGMCOE, Pune India*
Pooja Ingole*, computer department, SUUP,PGMCOE , Pune, India*
Rahul Ingole *, computer department, SSUP, PGMCOE, Pune, India*
Pooja Garje, computer department, SSUP, PGMCOE, Pune, India

was presented[7] to generate better discriminative features.

II.SYSTEM DESIGN

We are going to apply this mechanism on an online question portal website to detect DoS attack .It consists of three components first is Normal users who are going to register themselves and post the questions, second is administrator system in which detection mechanism is to be done using MCA and TMA, third is expert user who is going to answer the posted questions. Attackers may disrupt the system by posting number of meaningless questions and hence server may unable to provide service to its intended users.

III.SYSTEM ARCHITECTURE

In following section overview of our proposed DoS attack detection system is architecture is given and its framework with sample by sample detection mechanism is discussed.

A. framework

The detection mechanism can be divided into three phases as follows

(1) Phase one includes “Feature Normalization Module” in which basic information is generated from incoming network traffic to the internal network i.e. information about registered users. Protected servers are present at the internal network. Those servers are used to form visitors information specifically well defined time interval. As the monitoring and analyzing is done at destination network, this reduces the detection overhead[3]. This makes our detector to give best fit protection for the targeted network because the traffic profiles used by the detectors are developed for small number of network services.

(2)In the second phase basic information is collected from “Feature Normalization Module”. The multivariate correlate analysis is implemented. The triangle area map is generated which is used to extract the correlation between two distinct server records.The intrusive activities are identified by making them to cause changes to the correlation, with the help of these changes intrusions can be identified. All the obtained correlations are in the form of triangle areas which are then stored in Triangle Area Maps(TAMs).stored correlations are then used to replace the original basic features. Due to this better information is obtain which is useful to us to sort out the legitimate and illegitimate traffic records .

(3)In phase three the decision making is done using the anomaly based detection system. This gives information about any DoS attacks without the requirement of the relevant knowledge. The labor intensive attack analysis and misuse based detection are avoided. Two steps are involved in decision making (i.e. the training phase and test phase). The training phase consists of “Normal Profile Generation” which is used to generate profiles for various types of legitimate traffic records and these profiles are stored in the database. During the test phase the “Tested Profile Generation Module” builds profiles for individual traffic records which include both normal records and attack records. These profiles are then handed over to the attack detection module. This does the task of comparing the individual tested profile with respective stored normal profile. In attack detection module threshold –based classifier is used to distinguish the DoS attack from legitimate traffic.

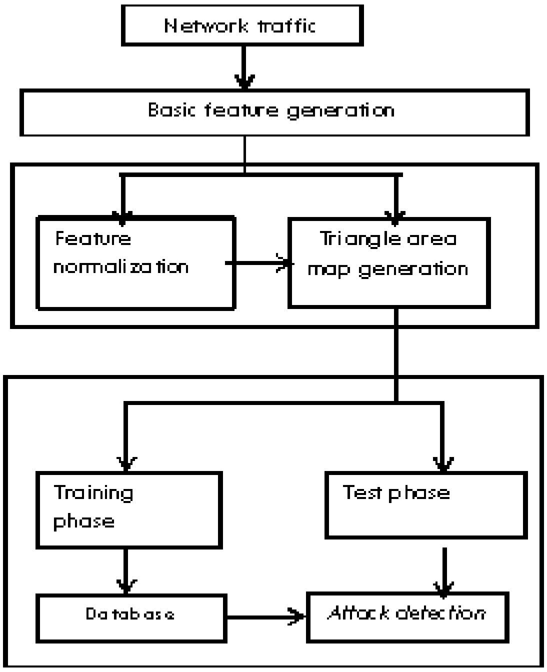

Fig – Denial Of Service Framework

B. sample by sample detection

It is systematically proved that the group-based detection mechanism maintained a higher probability in classifying a group of sequential network traffic samples than the sample-by sample mechanism. Whereas, the proof was based on an assumption that the samples in a tested group were all from the same distribution (class)[3]. This restricts the applications of the group based detection to limited scenarios, because attack occur unpredictably in general and it is difficult to obtain a group of sequential samples only from the same distribution. To remove this restriction, our system in this paper investigates traffic samples individually. This offers benefits that are not found in the group-based detection mechanism. For example, (a) attacks can be detected in a prompt manner in comparison with the group-based detection mechanism, (b) intrusive traffic samples can be labeled individually, and (c) the probability of correctly classifying a sample into its population is higher than the one achieved using the group-based detection mechanism in a general network scenario[3].There are following benefits of MCA technique, it does not require any relevant knowledge of network traffic.MCA withstand the peoblem of linear change in features.MCA provides characterization of individual traffic network.

IV.MULTIVARIATE CORRELATION ANALYSIS

By making use of the multivariate correlations, various types of network traffic can be clearly characterized. DoS attack traffic behaves differently from the legitimate network traffic, and the behavior of network traffic is reflected by its statistical properties. [4]To well describe these statistical properties, we present a novel Multivariate Correlation Analysis (MCA) approach in this section. This MCA approach employs triangle area for extracting the correlative information between the features within an observed data object. The details are presented in the following.

Given an arbitrary dataset X = {x_1, x_2, · · · , x_n},where x_i = [f_{i1} f_{i2}· · · f_{im}]^T , (1 ≤ i ≤ n) represents the i^{th} m-dimensional traffic record.

$$X = \begin{bmatrix} f_1^1 & f_2^1 & \cdots & f_m^1 \\ f_1^2 & f_2^2 & \cdots & f_m^2 \\ \vdots & \vdots & \ddots & \vdots \\ f_1^n & f_2^n & \cdots & f_m^n \end{bmatrix}$$

We apply the concept of triangle area to extract the geometrical correlation between the j^{th} and k^{th} features in the vector $x_{i.}$

$$x_i^T I = x_i^{'} = \begin{bmatrix} f_1^i & 0 & \cdots & 0 \\ 0 & f_2^i & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f_m^i \end{bmatrix}_{m\times m}$$

To obtain the triangle formed by the two features, data transformation is involved. The vector x_i is first projectedon the (j, k)th two-dimensional Euclidean subspace as $y_{i,j,k}$= [ε_i ε_k]^T x_i = [f_{jk} f_{ik}]^T, (1 ≤ i ≤ n,1 ≤ j ≤ m, 1 ≤ k ≤ m,j=k). The vectors ε_j= [$e_{i,1}$ $e_{j,2}$ · · · $e_{j,m}$]T and εk = [$e_{k,1}$ $e_{k,2}$ · · · $e_{k,m}$]^T have elements with values of zero, except the (j, j)th and (k, k)th elements whose values are ones in ε_j and ε_k respectively. The $y_{i,j,k}$ can be interpreted as a two dimensional column vector, which can also be defined as a point on the Cartesian coordinate system in the (j, k)th two-dimensional Euclidean subspace with coordinate (f_{ij} , f_{jk}). Then, on the Cartesian coordinate system, a triangle ΔfijOfik formed by the origin and the projected points of the coordinate (fij , fik) on the j-axis and k-axis is found. Its area Trij,k is defined as Tri j,k = ((fij , 0) − (0, 0) × (0, fik) − (0, 0))/2,(10)where 1 ≤ i ≤ n, 1 ≤ j ≤ m, 1 ≤ k ≤ m andj = k. In order to make a complete analysis, all possible permutations of any two distinct features in the vector x_i are extracted and the corresponding triangle areas are computed. A Triangle Area Map (TAM) is constructed and all the triangle areas are arranged on the map with respect to their indexes. For example, the $Tr^i_{j,k}$ is positioned on the j^{th} row and the k^{th} column of the map TAM^i, which has a size of m × m. The values of the elements on the diagonal of the map are set to zeros ($Tr^i_{j,k}$ = 0, if j = k) because we only care about the correlation between each pair of distinct features. For the non-diagonal elements $Tr^i_{j,k}$ and $Tr^i_{k,j}$ where j = k, they indeed represent the areas of the same triangle.This infers that the values of $Tr^i_{j,k}$ and $Tr^i_{k,j}$ are actually equal. Hence, the TAM^i is a symmetric matrix having elements of zero on the main diagonal.When comparing two TAMs, we can imagine them as two images symmetric along their main diagonals. Any differences, identified on the upper triangles of the images, can be found on their lower triangles as well.Therefore, to perform a quick comparison of the two TAMs, we can choose to investigate either the upper triangles or the lower triangles of the TAMs only. This produces the same result as comparing using the entire TAMs (see Appendix 1 in the supplemental file to this paper for an example). Therefore, the correlations residing in a traffic record (vector x_i) can be represented effectively and correctly by the upper triangle or the lower triangle of the respective TAM^i For consistency, we consider the lower triangles of TAMs in the following sections. The lower triangle of the TAM^i is converted into a new correlation vector TAM^i $_{lower}$ denoted as TAM^i $_{lower}$ = [$Tr^i_{2,1}$ $Tr^i_{3,1}$· · · $Tr^i_{m,1}$ $Tr^i_{3,2}$ s

$Tr^i_{4,2}$· · · $Tr^i_{m,2}$· · Tr^i $_{m,m-1}$]^T .

For the aforementioned dataset *X*, its geometrical multivariate correlations can be represented by $XTAM_{lower}$ = {TAM^1_{lower}, TAM^2_{lower}, · · · , TAM^i_{lower}, · · · , TAM^n_{lower}}.When putting into practice, the computation of the*Trij,k* defined in can be simplified because the value of the *Trij,k* is eventually equal to half of the multiplication of the absolute values of *fi j* and *fik*. Therefore,the transformation can be eliminated, and can be replaced by *Trij,k* = (|*fij*| × |*fik*|)/2. The above explanation shows that our MCA approach supplies with the following benefits to data analysis. First, it does not require the knowledge of historic traffic in performing analysis. Second, unlike the Covariance matrix approaches proposed in which is vulnerable to linear change of all features, our proposed triangle-area-based MCA withstands the problem. Third, it provides characterization for individual network traffic records rather than model network traffic behavior of a group of network traffic records. This results in lower latency in decision making and enables sample-by-sample detection. Fourth, the correlations between distinct pairs of features are revealed through the geometrical structure analysis. Changes of these structures may occur when anomaly behaviors appear in the network. This provides an important signal to trigger an alert.[6]A Triangle Area Map (TAM) is constructed and all the triangle areas are arranged on the map with respect to their indexes. Hence, the TAM^i is a symmetric matrix having elements of zero on the main diagonal .The two primary advantages of the proposed analysis technique are supported by two underlying mathematical structures. They are the transformed traffic record matrix and the Euclidean distance. These two mathematical tools help solve the dilemmas caused by the occurrence of two distinct pairs of features having the same distance on one-dimensional space and the linear change of all features.

V.DETECTION MECHANISM

In this section, we present a threshold-based anomaly detector. This detector generates normal profiles using purely legitimate network traffic records and uses these records for future comparisons with new incoming investigated traffic records. The dissimilarity between a new incoming traffic record and the respective normal profile is examined by the proposed detector[5]. If the dissimilarity is greater than a pre-determined threshold, the traffic record is considered as an attack. Otherwise, it is labeled as a legitimate traffic record. Performance of a threshold-based detector is depends upon normal profiles and thresholds.[1] 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, The triangle-area based MCA approach is applied to analyze the records. [1] 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. Unlike Euclidean distance and Manhattan distance, it evaluates distance between two multivariate data. objects by taking the correlations between variables into account removing the dependency on the scale of measurement during the calculation. Finally, the obtained distribution of the normal training traffic records, are stored in the normal profile for attack detection.

B. Threshold Selection

Threshold = $\mu + \sigma * \alpha$.

For a normal distribution, α is usually ranged from 1 to 3.The threshold given is used to differentiate attack traffic from the legitimate one. if the MD between an observed traffic record and the respective normal profile is greater than the threshold, it will be considered as an attack. Many threshold frequency were set in comparison. The result reveals that at a certain threshold the server goes to sleep mode for long time period and crashes. Now this particular threshold is set as a limit to detect the intrusive networks.

C. Attack detection

To detect DoS attacks, the lower triangle (TAM observed lower) of the TAM of an observed record needs to be generated using the proposed triangle-area-based MCA approach[6]. Then, the MD between the TAM observed lower and the TAM normal lower stored in the respective pre generated normal profile Pro is computed using the detailed detection algorithms.

VI.COMPARISON WITH DIFFERENT DETECTION APPROACHES

	Triangle area based nearest neighbors approach	Euclidean distance map based approach original data]	The proposed detection system [original data]	The proposed detection system [normalized data]
Accuracy	92.15%	99.87%	95.20%	99.95%

VI. CONCLUSION

. This paper has offered a MCA-based totally DoS attack detection gadget that's powered by the triangle-area based MCA approach and the anomaly-based(behavior based) detection method. the former method extracts the geometrical correlations hidden in pairs of features inside every network visitors record, and gives more accurate characterization for community traffic behaviors. The latter approach helps our system in order to distinguish each acknowledged and unknown DoS attacks. assessment has been carried out using [2]KDD Cup 99 dataset to affirm the effectiveness and overall performance of the proposed DoS assault detection system. The impact of original (non-normalized) and normalized records has been studied within the paper. The outcomes have discovered that after operating with non-normalized statistics, our detection machine achieves maximum ninety five. The problem, but, can be solved through utilizing statistical normalization technique to remove the bias from the facts. The consequences of comparing with the normalized records have shown a greater encouraging detection accuracy of ninety nine. ninety five% and almost a hundred .Besides, the comparison end result has tested that our detection gadget outperforms. moreover, the computational complexity and the time price of the proposed detection machine have been analyzed . To be a part of the future paintings, we are able to in addition check our DoS attack detection device the use of real world records.

ACKNOWLEDGMENT

First and foremost, I would like to thank our guide Prof.Shrikant Dhamdhere for his guidance and support.We will forever remain grateful for the constant support and guidance extended by guide, in making of the report. Through our many discussions, he helped us to form and solidify ideas. The invaluable discussions I had with him, the penetrating questions he has put tomes and the constant motivation, has all led to the development of this project.

We would like to express our deepest appreciation towards our Principal and Prof. Shrikant Dhamdhere, Head of the Department whose invaluable guidance supported us in completing this project.

We are equally indebted to all faculty members of

Computer Department. This project work has been greatly assisted by the co-operation of Computer Laboratory Staff and Library Staff who provided kind support and facilities.

REFERENCES

[1] V. Paxson, “Bro: A System for Detecting Network Intruders in Realtime,”Computer Networks, vol. 31, pp. 2435-2463, 1999

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

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

[4] G. Thatte, U. Mitra, and J. Heidemann, “Parametric Methods for Anomaly Detection in Aggregate Traffic,” Networking, IEEE/ACM Transactions on, vol. 19, no. 2, pp. 512-525, 2011.

[5] Z. Tan, A. Jamdagni, X. He, P. Nanda, and R. P. Liu, “Denialof-Service Attack Detection Based on Multivariate Correlation Analysis,” Neural Information Processing, 2011, pp. 756-765.

[6]. A System for Denial-of-Service Attack Detection Based on Multivariate Correlation
Analysis Zhiyuan Tan, Aruna Jamdagni, Xiangjian He‡, Senior Member, IEEE,

[7] C. F. Tsai and C. Y. Lin, “A Triangle Area Based Nearest Neighbors Approach to Intrusion Detection,” Pattern Recognition, vol. 43, pp.222-229, 2010.

Ankush Bhat
B.E. PGMCOE,SPPU,Pune,India

Pooja Igole
B.E. PGMCOE,SPPU,Pune,India

Rahul Ingole
B.E. PGMCOE,SPPU,Pune,India

Pooja Garje
B.E. PGMCOE,SPPU,Pune,India