EARTH MOVER DISTANCE SYSTEM TO DISCLOSE DOS ATTACKS

Nitin A Varghese[1], Rachana Rai[2], Smitha H L[3], Tanya Priya[4], Mahesh T R[5]

[1,2,3,4]*Department of Computer Science and Engineering, T. John Institute of Technology, Bengaluru, India*
nitinvarghese829@gmail.com, rachurai19@gmail.com, hlsmithahl@gmail.com, tanyapriy@gmail.com
[5]*Head of the Department, Department of Computer Science and Engineering, T. John Institute of Technology, Bengaluru, India*
mahesh@tjohngroup.com

Abstract

In this internet era, security is a major issue. Out of the many attacks DoS attack is one among them. Researchers have been working on it from 1990's. Many systems were introduced to detect DoS attacks which used machine learning approaches and statistical analysis where in, in the proposed system we use anomaly based technique. The methods used for detection are Principal Component Analysis, Multivariate Correlation Analysis using Triangular Area Map and Earth Mover Distance. For reducing the features we make use of the Principal Component Analysis technique which is the dimensionality reduction algorithm. For effective detection , finding the correlation between the obtained features is important and hence we use Multivariate Correlation Analysis along with the Triangular Area Map. Out of the many known dissimilarity measures in MinKowski-form distance L_P and X^2 statistics where it evaluates dissimilarity between distributions, we make use of different approache called Earth Mover Distanc(EMD) to attain the goal which gives higher accuracy. The uniqueness of EMD makes the proposed system more capable. Evaluations are conducted using KDD cup dataset.

Keywords: *Denial of Service, Principal Component Analysis, Multivariate Correlation Analysis, Triangular Area Map, Earth Mover Distance.*

---***---

I. INTRODUCTION

The growing usage of internet poses many threats for the security of the data. The world encounters many types of attacks for example active attacks, passive attacks, distributed attacks etc. In that Denial of Service attack is a kind of an attack we deal with.

Denial of service attack occurs when an attacker tries to make the resources or machine partly or completely unavailable to the user when the system is asked for a particular service. It basically blocks the server from acknowledging the services when it is asked for.

The DoS attacks can be launched in the systems across the network deliberately which can cause serious issues to the victim, can even exploit loop holes in the system which are usually referred as system vulnerabilities. Usability to provide the required service at that particular time can lead to financial crisis for service providers. Attackers can attack the system in various possible ways. Firstly, by flooding a victim with too many requests which will block the server to provide the services requested hence, it may lead to system crash.

Secondly, the attacker can catch hold of the designated resources like memory, processor time, external devices which will affect the fulfillment of other services.

Lastly, in this modern world, attack toolkits are freely and readily available and even if the attackers has less knowledge about the network security can easily make use of these tools to successfully make an attack.

There are four strategies to oppose DoS attacks and they are Detection, Prevention, Mitigation and Response, where in, in our system, the key concern is on Detection, which is the first and foremost step to defend against an attack.

The detection mechanism is classified into two types-misuse based and anomaly based detection. Misuse based Detection- Even though it showcases high rate in detecting an attack this approach suffers because it cannot find the attacks which have slight variations or deviations in the behaviors. Anomaly based Detection- In this mechanism, even the smallest variations seen in the behaviors of the profiles are considered to check if the request is a normal request or an attack. To gain high accuracy in the proposed system we use Anomaly based Detection mechanisms. The key feature of this proposed system is that it uses Earth Mover Distance mechanism which provides greatest efficiency till date.

II. RELATED WORK

Existing systems mostly use misuse based detection techniques and anomaly based detection techniques. Previously systems were using machine learning and statistical analysis, however they face issues regarding attaining high accuracy when detecting both normal requests and attacks

Some of the previously existing systems use Mahalanobis distance to disclose an attack where in, in the existing system the records were detected based on Mahalanobis distance to find distance between the network distribution and client point.

All the data features obtained from the network is processed using Multivariate Correlation Analysis to determine if it is a normal request or an attack.

In case of misuse based detection where signature matching technique is used, it is very clumsy and labor intensive task to generate the signatures for previously unseen attacks. The disadvantage of this technology is that it is not possible to recognize hidden individual attack records in the given set.

III. PROPOSED SYSTEM

Unlike formerly based detection system, this system uses Anomaly based detection mechanism which uses rule matching concept to classify whether the request is normal or anomalous. Misuse Based Detection can attend high accuracy only in the case of known attacks wherein anomaly based detection can detect attacks within a range. If the client vary the features even system can change and detect the attack efficiently. Mechanism to detect the request studies the relation between all the features of the request sent by the client end. Records of the individual attacks which can be unseen in a sample set can be easily revealed.

Proposed system possess three main key features they are

The first and foremost one is **Principle Component Analysis (PCA)** technique. A dataset is given as an input to it and the technique performs the conversion of the distributed data to a unidirectional data. Thus dimensions of data are reduced. Eigen vector is generated and this technique expertise the system in gaining accuracy.

```
Requirement: Data set Y {Y contains a instances , and each of which has b features}
Ensure: 1 ≤ n ≤ b
$\bar{y} \leftarrow \frac{1}{a}\sum_{j=1}^{a} y_j$
$Yzm \leftarrow Y - \bar{y}$ {Subtract $\bar{y}$ from each instances in Y}
$K_Y \leftarrow \frac{1}{a-1} Y_{zm} Y_{zm}^T$
Obtain M and E which are subjected to $ME = K_Y E$
for $j = 1$ to $a$ do
    $\theta_j^2 = \sum_{b=1}^{j} \phi_b$
end for
Plot { $\theta_1^2, \theta_2^2, ....., \theta_n^2$ }
Locate the "elbow" on the scree plot and identify the index(n) of "elbow" point.
$E_n \leftarrow$ The selected first n eigen vector of E.
return $E_n$
```

Algorithm for PCA

Next feature specifies about the **Multivariate Correlation Analysis (MCA)**. Eigen vector produced above is fed as an input to this method. The technique plays a major role in finding the relations between the features of the request sent by the client. Triangular Area Map is used along with Multivariate Correlation Analysis to speed up the computation and analyze the correlations swiftly.

```
Requirement: Data set $Y$ and subspace $E_n$ {$Y$ contains a instances, and each of which has b features. $E_n$ is the selected first n eigenvectors of $E$ }
Initialize RAN {It is an array with a element denoted by $Ran_j (1 \le j \le a)$}
Initialize $Y_{TRI}$ with a n-by-n matrices denoted by $TRI^i (1 \le j \le a)$
$Y_{GN} \leftarrow Y \times E_n$ { $Y_{GN}$ contains a instances, and each of which has n features}
for $j = 1$ to $n$ do
    $TRI^j \leftarrow [Tm^j_{i,q}]_{n \times n}$ ,where $1 \le i, q \le n$
    {Triangle area formed involving the features i and q of $Y_{GN}$ is computed and assigned to the (i,q)-th element in $TRI^j$}
end for
$\overline{TRI} \leftarrow \frac{1}{a}\sum_{j=1}^{a} TRI^j$
for $j = 1$ to $n$ do
    $Ran_j \leftarrow EMD - L_1(TRI^j, \overline{TRI})$
    {Earth mover's distance between $TRI^j and\ \overline{TRI}$ }
end for
$\overline{RAN} \leftarrow \frac{1}{a}\sum_{j=1}^{a} RAN_j$
$Dev = \sqrt{\frac{1}{a}\sum_{j=1}^{a}(Ran_j - \overline{RAN})^2}$
$Gen \leftarrow (\overline{TRI}, \overline{RAN}, Dev)$
return $Gen$
```

Algorithm based on MCA

Last one depicts about **Earth Mover Distance (EMD)**. EMD technique gives the interval between the point and the distribution where point can be considered as client and distribution as KDD cup dataset over a region. Usually detection mechanism of Denial of Service attack uses Mahalanobis distance to find out the difference whereas this particular system uses Earth Mover Distance which ensures greater proficiency.

```
Requirement: Tested sample $y_{test}$ ,subspace $E_n$ , normal profile generated $Gen$ and parameter $\beta$
$y_{test}^{GN} \leftarrow y_{test} \times E_n$ {Project tested sample $y_{test}$ onto
    the subspace $E_n$ }
$TRI_{test} \leftarrow [Tm_{i,q}^{j}]_{n\times n,}$ ,where $1 \le i, q \le n$
$Ran_{test} \leftarrow EMD - L_1(TRI_{test}, \overline{TRI})$
if $(\overline{RAN} - \beta \times Dev) \le Ran_{test} \le (\overline{RAN} + \beta \times Dev)$
  then
        return Normal
else
        return Attack
end if
```

Algorithm for EMD

System consists of three modules:-

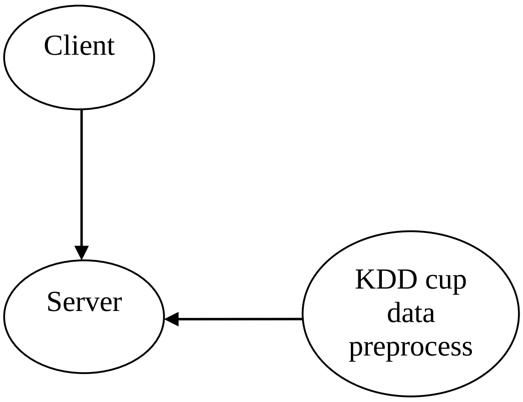

Fig 1: Interaction between the modules

1. **KDD cup pre-processor**: Client is someone which can access only those services which the server provides them. Therefore if they need any resource or service, it has to request to the server for the same. Request which client send can be of any kind normal request or an attack. Hence the server has to be vigilant regarding the request it encounters so that an attack cannot affect him in any possible way. Server maintains a track of the entire request which comes to it and establishes a profile.

   For the generation of profile, KDD first undergo Principle Component Analysis technique which takes KDD cup dataset which consist of the features of Denial of Service attacks .Evaluation of Eigen vector take place.

   Next Multivariate Correlation is computed by multiplying the attribute values of the features to the Eigen vector obtained. Then Triangular Area Map comes into the picture in which two triangles are obtained known as upper triangle and lower triangle but then we can consider only one triangle since the values of both the triangle remain same and this consideration makes the evaluation smaller and computation fast. Earth Mover Distance is the one which calculate the gap between the KDD cup dataset and the client.

   Finally a profile for the features that were given is estimated. In the end based on the profile we come to a conclusion that request of the client is genuine or not.

2. **Client**: Client should be registered to the server before it requests a server. An unregistered client tries to access the services is considered as illegal and can be interpreted as an attacker. So even in this model initially a client has to register to the server. Authentication checking will take place in server side so that server gets confirmation about the user. If authentication is successful server send a onetime password to the client in his mail. Now the client can send the features to the server. If the authentication fails process stops their and client cannot send its features the server and accordingly they cannot ask for resource or anything to the server.

3. **Server**: Server is the one which fulfil all the demands of the client. Formerly system employs Principle Component Analysis technique which is performed to decrease the dimensionality of the data where the data can be features related to the attack or the features related to the request. While performing the technique Eigen vector is reproduced this is given as an input to the next step. Then the relation between the features obtained is correlated using Multivariate Correlation Analysis mechanism along with triangular area map. The Eigen vector got from the previous step is multiplied with the features of different attacks and finally correlation analysis is created by projecting these values into Triangular Area Map. Lastly when the Triangular Area Map values are obtained, dataset becomes normalized. Further Earth Mover Distance is used for measuring the dissimilarity between the upcoming traffic and prebuilt profile. And hence the profile created in KDD cup module is compared to find out whether the request is normal or an attack. Earth Mover Distance takes mean of Triangular Area Map produced from KDD cup side. Range in which an attack can be detected is represented in the Threshold Selection .It takes the input as standard deviation and mean of Earth Mover Distance coming up from the KDD cup pre processor end.

IV. DESIGN

Client requirement is fulfilled by server. Any service or resource which the client needs it has to request to the server. Now the scenario is that there can be any number of requests a client can upload to the server at an instance. Sometimes client is rigid to get any resource by hook or crook, if not legally then illegally and this mentality of the client can ruin the server. Therefore the server has to be alert with respect to the clever clients. For that the server should maintain the record of the entire request encountered from the client. Many requests in general is termed as traffic. Framework of the proposed system initially depicts the normal traffic which the client sends to the server.

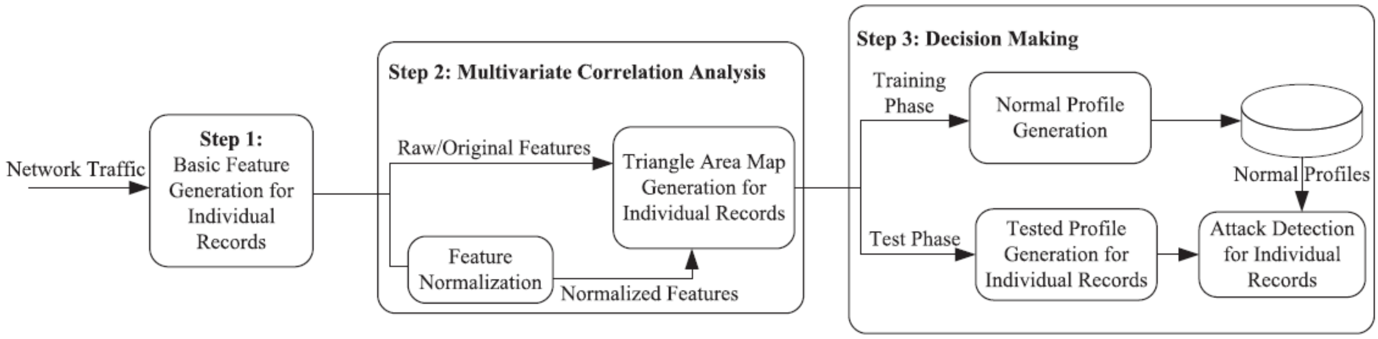
Network Traffic
Step 1:
Basic Feature Generation for Individual Records
Step 2: Multivariate Correlation Analysis
Raw/Original Features
Triangle Area Map Generation for Individual Records
Feature Normalization
Normalized Features
Step 3: Decision Making
Training Phase
Normal Profile Generation
Normal Profiles
Test Phase
Tested Profile Generation for Individual Records
Attack Detection for Individual Records

Fig 2: System architecture

Step 1: Server is not concerned about the client personally it is just concerned of the features of the request the client has sent. These features are used to produce individual records for each and every request that the client sends to make sure that request made is not an attack.

Step 2: Principal Component Analysis technique is computed where the data which is distributed is made to normalized. Next Multivariate Correlation analysis is performed wherein original features sent from the client end and the normalized data is multiplied to find the relationship between the features. Original features and normalized data are fed as an input to the Triangular area map which accomplishes individual records generated to speeds up the computation of Triangular Area Map.
Lastly the difference is evaluated between the network and client over an area.

Step 3: The third step indicates Decision making, an area which specifies the server that request is a normal request or an attack. This step includes training phase and test phase. In the training phase PCA, MCA and EMD is performed on KDD cup data set and a profile is generated

Where as in the test phase the client features that is extracted by the server undergoes PCA, MCA and EMD and again a profile is generated. Then both the profiles are matched with each other to detect whether the request sent was a genuine or an attack.

V. CONCLUSION

In this paper we are discussing about Denial of Service attacks detection system which consist of Principle Component analysis, Multivariate corelation analysis and Earth mover's distance. MCA technique helps to find the corelation among each featuer of different client request that comes in the traffic and gives exact improvisation for network traffic understanding. The later mechanism allow a system to easily identify the known and unkown DoS attack from appropriate network traffic.

Calculation for the attack in the traffic is done using KDD cup dataset.The result tells that our detection system discloses the attack with a higher rate as compared to all the existing system so far.Our proposed system's computational compexity can be easily comapred with the state-of- art approaches.By the test that we have done on the system it is easily understood that our detection sysytem can work in high speed network segments.

REFERENCES

[1] V. Paxson, "Bro: A System for Detecting Network Intruders in Realtime,"Computer Networks, vol. 31, pp. 2435-2463, 2006.
[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] D. E. Denning, "An Intrusion-detection Model," IEEE Transactions on Software Engineering, pp. 222-232, 2010.
[4] K. Lee, J. Kim, K. H. Kwon, Y. Han, and S. Kim, "DDoS attack detection method using cluster analysis," Expert Systems with Applications, vol. 34, no. 3, pp. 1659-1665, 2010.
[5] A. Tajbakhsh, M. Rahmati, and A. Mirzaei, "Intrusion detection using fuzzy association rules," Applied Soft Computing, vol. 9, no. 2, pp. 462-469, 2011.
[6]W. Hu, W. Hu, and S. Maybank, "AdaBoost-Based Algorithm for Network Intrusion Detection," Trans. Sys. Man Cyber. Part B, vol. 38, no. 2, pp. 577-583, 2012.
[7] 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, 2013.
[8] C. Yu, H. Kai, and K. Wei-Shinn, "Collaborative Detection of DDoS Attacks over Multiple Network Domains," Parallel and Distributed Systems, IEEE Transactions on, vol. 18, pp. 1649-1662, 2013.
[9] W. Wang, X. Zhang, S. Gombault, and S. J. Knapskog, "Attribute Normalization in Network Intrusion Detection," The 10th International Symposium on Pervasive Systems, Algorithms, and Networks (ISPAN), 2015.
[10] W. Wang, X. Zhang, S. Gombault, and S. J. Knapskog," A system for detecting Denial of Service using multivariate correlation analysis"Networking ,IEEE Transactions on, vol. 32 , no.5,pp. 1025-1032,2015.