Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IMPACT FACTOR: 6.017

IJCSMC, Vol. 5, Issue. 11, November 2016, pg.77 – 81

An Anomaly Based Detection System for DOS Attack Using Correlation Analysis

Mr. Krishnaraj, Mrs. Asmita Poojari

Department of Computer Science and Engineering, NMAMIT, Nitte, India

Department of Computer Science and Engineering, NMAMIT, Nitte, India

Abstract— ***Most of the organizations, enterprises will provide the services through online to achieve faster delivery of their services. The clients or user of the organization will access the registered services through internet. In such cases, users may be delayed or denied from accessing their services although they are authorized to access the services because of DOS attack. To avoid DOS attack on different servers such as web servers, database server etc., an anomaly based detection system is proposed that uses correlation between the incoming streams. It uses the ideas of correlation analysis and Mahalanobis Distance to detect the attacks. Correlation analysis is required to know the relationship among the different features of the incoming streams. Mahalanobis distance method is used to know the dissimilarity between the legitimate and illegitimate network streams. This system has the ability to identify both known as well as unknown attacks.***

Keywords— ***Internet, Correlation Analysis, DOS, Online service, Mahalanobis Distance***

I. Introduction

Denial of Service attack is one of the serious attacks to the online servers now a day. This type of attack causes changes to the behaviour of the servers. DOS attack degrades the performance of the entire system by making the host, router or any other component of the network, as unavailable. This can be caused by flooding the server or network component with large amount of unnecessary packets that are not useful. The functionality of the server or other network component can be stopped for few minute or for few days. This really impacts on the organization and damages the services that are offered by the organization. Hence, for the effective functionality of the system, it is required to detect such attacks and need to provide the protection to the online servers. Network detection system can be grouped in two groups called misuse based network detection system and anomaly based network detection systems. Misuse based detection systems identify the attack by observing the network traffic continuously and by matching them with the known attack signatures. This kind of detection system will not identify any new attacks and even variants of the known attacks. Creating and updating the attack signature is a manual process. Hence, it becomes expensive to keep the signature database up to date due to labour cost. In Anomaly based detection systems, detection process is not based on the attack signature. The profiles of the legitimate network traffic are obtained by using data mining algorithms and statistical analysis. This is not a manual process and hence do not need network experts to generate the attack signature

II. Purpose

This paper is aimed towards finding an algorithm for DOS attack detection using Multivariate correlation analysis. Attack detection mechanism should be able to process the individual packets in order to determine the correlation among the packets. This is required because, an intruder may include the illegitimate information in more than one packets. Correlation analysis approach allows revealing such illegitimate information from different packets and detecting the attack.

A. Existing System

Different algorithms have been proposed to detect DOS attacks for online servers. All detection system will manipulate group of packets from different sources and concludes all group of packets either as legitimate stream or illegitimate stream using the detection mechanism. Moreover, they require the prior knowledge of attacks to determine the attack traffic. These mechanisms cannot identify any new type of attack and cannot process the network traffic individually

B. Proposed System

The proposed system is developed to detect DOS attack for online servers and it is based on multivariate correlation analysis approach. It has the ability to process the incoming network traffic packet by packet. Also, it does not require any prior knowledge regarding the attacks and hence it detects all type of new attacks. Basic features are identified by using data analysis and correlations among these features are analysed using multivariate correlation analysis approach. Normal profiles will be generated from these analyzed features and they are stored in the database and this is considered as the normal profile for legitimate traffic. After generating the normal profiles, all the incoming network traffic is analyzed using the same procedure and normal profile generated is compared with the normal profile of the legitimate traffic. The detection process is designed based on threshold value. The incoming traffic is considered as illegitimate traffic, if its normal profile varies more than the predefined threshold value and considered as DOS attack.

III. LITERATURE SERVEY

This chapter provides the information about all the past works that are done in this area. S. Yu et al [1] described a method to overcome DOS attack by analyzing the correlation among the suspicious flow and normal flow. He proposed an algorithm to detect the attack by considering the flow coefficient as the factor to identify the dissimilarity between the suspicious and normal flow. Tsai et al [2] proposed an algorithm to detect the DOS attack using nearest neighbours method based on triangle area approach. Other works that are carried out in this field are as follows.

A. Jamadagni et al [3] presented an intrusion detection system based on real time payload in a multitier architecture. This paper explains how to give protection to the network component in real time environment. Z. Tan et al [4] presented a system that efficiently identifies DOS attacks using Multivariate correlation analysis. This paper explains how to analyse the correlation between the incoming streams. Z. Tan et al [5] presented an idea to detect the DOS attacks based on triangle area and correlation analysis approach. This paper describes the method of detecting the attack using triangle area. J. S. Baras et al [6] have presented an algorithm to detect worms, DOS attack and to detect other types of network attacks in a distributed environment. This paper explains detecting the network attack in distributed networks. W. Wrang et al [7] presented a method to normalize the attributes in order to detect network intrusions. This paper describes the method to normalize the network attributes and then how these attributes can be used to detect the intrusion. P. Garca et al [8] have conducted a study on anomaly based systems and challenges in anomaly based intrusion detection system. It also provides technics to detect anomaly based network attacks. V. Paxson et al [9] presented a paper on detecting network intruder in real time. It explains a mechanism to identify the attacker in real time. S.J. Stalfo et al [10] described a method to identify the basic network features from the incoming stream. It specifies that, basic features can be obtained by analyzing the streams and by using data mining algorithms. K. Lee et al [11] presented a detection model to identify the distributed DOS attacks by analyzing the clusters. This paper also explains the steps to be taken to analyse the clusters.

IV. SYSTEM ARCHITECTURE

The proposed system of DOS attack detection is shown in Fig 1. This system contains four important modules. They are, Feature Generation, Multivariate Correlation Analysis, Training Phase, and Decision Making.

A. Feature Generation

In this module, features or attributes are identified from the incoming stream. Features are nothing but the attributes that are obtained by analyzing the behaviour of the network traffic. These attributes can be obtained by applying any data mining algorithms over the large set of collected data from the user from the last few years. These features will provide the idea about the attributes that are used frequently by the attacker. For example, in case of bank server, to communicate with server, the user may provide information such as user name, user id, password, account number, debit card number etc. But, the system will consider only such attributes that are being used most of the time by the attacker to attack the server.

B. Multivariate Correlation Analysis

This is the technic adapted to know the correlation among the different packets. In this phase, the system will consider only the features that are obtained in the feature generation step. Consider, the data set X = $\{f_1, f_2, f_3 \ldots . f_n\}$ represents the set of basic features obtained in the previous step. This analysis uses triangle area to get the relationship among the features. The triangle area between any two features can be obtained using the formula,

$$Tr_{j,k}=(|f_j| * |f_k|) / 2.$$

This gives the correlation between the features ‘f_j’ and ‘f_k’. To analyse all the correlation among all the features, it is needed to calculate the areas of all the combinations of any two features. After all the areas are calculated, it is needed to represent them in a two dimensional matrix. Mapping the areas into two dimensional matrixes can be done based on their indexes. For example, the area of the two features f_2 and f_3 can be mapped into the location (2, 3) in the matrix. The diagonal values of this matrix will be set to all zeros. Moreover, it is not required to consider all the entries of the matrix since the values above and below the main diagonal are identical. So, we need to consider either the upper diagonal values or lower diagonal values. This provides the relationship among all the features of the data set.

C. Training Phase

The system will work in two different modes namely training mode and test mode. In training mode, the system will be trained for all the different types of legitimate stream and corresponding profiles will be stored in the database. These normal profiles are considered in attack detection for the illegitimate stream. Whenever a new service is added to the server, the system will enter into the training mode to generate the profile for that new service. In test mode, all the incoming traffic is processed to determine whether it is an attack or not. The same procedure is carried out as that of the training mode and profiles are generated. Generation of normal profile for new type of services is carried out automatically without the intervention of network experts.

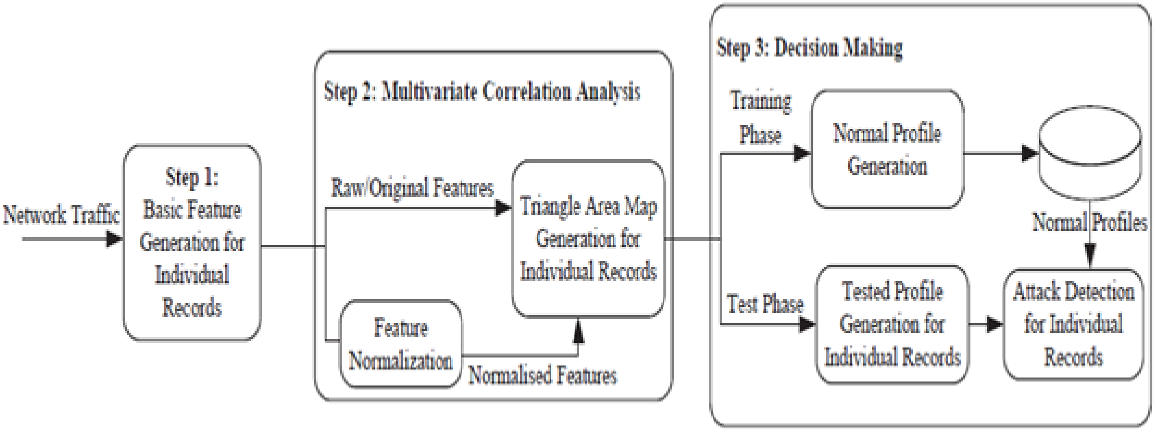
Network Traffic
Step 1:
Basic Feature Generation for Individual Records
Step 2: Multivariate Correlation Analysis
Raw/Original Features
Triangle Area Map Generation for Individual Records
Feature Normalization
Normalised Features
Step 3: Decision Making
Training Phase
Normal Profile Generation
Normal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

Fig 1: System Architecture of DOS attack detection System

D. Decision Making

This is the important part of the system and it is responsible for identifying the attack. The attack will be detected based on the threshold value. For all the incoming traffic normal profiles are generated and these normal profiles are compared with the normal profiles of the legitimate traffic that are already stored in the database. If the observed normal profile is above the predefined threshold, decision making module will consider that traffic as illegitimate traffic and corresponding action can be taken. The value of the threshold is important and can be fixed statically or dynamically as follows.

$$\text{Threshold}= \mu + \sigma * \alpha$$

Where ‘μ’ is mean and ‘σ’ is standard deviation. The value of ‘α’ is usually ragged from 1 to 3 for a normal distribution.

V. IMPLEMETATION

A. Pseudo Code for Correlation Analysis:

This module is used to characterize the behavioural differences between the legitimate and illegitimate network traffic. To know these behavioural differences it is required to understand the relation among the packets. This module accepts set of data packets as inputs and generates set of mapped correlated values in a vector for the given data set. Below code describes the actual mechanism involved.

```
Input: A data set X= { $f_1$, $f_2$, $f_3$….$f_n$}
Output: n x n matrix with correlated mapped values.
    1. Represent all the features of data set in ‘n’ dimensional matrix.
    2. FOR every data set
           FOR every two combination of features
               Calculate area using $|f_j|$ x $|f_k|$ /2
               Map the result into n x n matrix
           END FOR
       END FOR
    3. FOR every data set
       Consider only the upper or lower diagonal values of the matrix.
       Represent them in one dimensional matrix.
       END FOR
```

Fig 2: Correlation Analysis

B. Pseudo Code for Profile Generation:

As it is mentioned before, this system has two modes namely training mode and test mode. In the training mode, profile is generated for the legitimate stream and stored in the database. This is the profile generation module that takes the matrix of mapped correlated values and generates normal profiles in terms of mean and standard deviation. The steps are explained as below.

```
Input: n x n matrix with correlated mapped values.
 Output: Normal profiles for the incoming stream.
    1. Compute expectation of ‘g’ legitimate traffic record using,
              1/g $\sum_{i=1}^{g} TAM^{normal,i}{}_{lower}$
    2. Find out covariance matrix for $X_{TAM}{}^{norma}{}_{lower}$ .
    3. FOR i=1 to ‘g’ do
              Calculate Mahalanobis distance from legitimate traffic and expectation.
       END FOR
    4. Represent MDs using two parameters: mean and standard deviation.
    5. Store obtained distribution of normal training record, expectation and covariance matrix in the
       database.
```

Fig 3: Profile Generation

C. Pseudo Code for Attack Detection:

This module is used to determine whether the incoming stream is either a legitimate or illegitimate stream. In the training mode of the system, profiles are generated and stored in the database. In the test mode, every incoming stream is passed through the same procedure and profiles are generated. This generated profile and stored profiles are compared and if they vary beyond the threshold, incoming stream is considered as attack stream. This module requires profiles of the legitimate stream for its operation.

```
Required: Normal Profiles (distribution of normal training record, expectation and standard deviation)
  1.  FOR every incoming stream
              Calculate area for all combinations of features
              Represent them in n x n matrix.
      END FOR
  2.  Calculate MD between observed incoming stream and the respective normal   profile.
  3.  Fix the threshold.
  4.  IF (observed MD is greater than the threshold)
              Return ATTACK
      Else
              Return NORMAL.
      END IF
```

Fig 4: Attack Detection

VI. CONCLUSION AND FUTURE WORK

This system is designed to detect the DOS attacks based on correlation analysis. To detect the attack accurately, it uses Mahalanobis Distance method which easily identify the dissimilarity among the different network stream. First, basic features are identified those can be obtained by analyzing the large user data by using data mining algorithms. Then, correlation among these features is calculated and normal profiles is generated and stored in the database. Based on the threshold, we are determining the attack and after detecting the attack we can take any corrective actions such as blocking the user.

This detection system has the ability to detect both known and unknown attack since it does not need any previous knowledge regarding the attacks. This system will consider every input data as normalized data. Whenever non-normalized input data is provided to the system, it may not be able to detect the DOS attacks efficiently. So, in future we can implement any normalization methods to process non-normalized input data before applying DOS attack algorithm. Again, this system will process data only of text format. It will not process the features of incoming stream if the stream contains data in the form of images, audio etc.

REFERENCES

[1] S.Yu, W.Zhou, W.Jia, S.Gio, Y.Xiang and F.Tang, “Discriminating DDoS Attacks from Flash Crowds Using Flow Correlation Coefficient” Parallel and Distributed Systems, IEEE Transactions on, vol 23,pp. 1073-1080,2012.

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

[3] A. Jamdagni, Z.Tan, X. He, P. Nanda, and R.P. Liu, “RePIDS: A multi-tier Real-time Payload-based Intrusion Detection System,” Computer Network, vol. 57, pp. 811-824, 2013.

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

[5] 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,” The 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications, Livepool, United Kingdom, 2012, pp. 33-40.

[6] A.A. Cardenas, J.S. Baras, and V.Ramezani, “Distributed Change detection for worms, DDoS and other network attacks,” The American Control Conference, Vol.2, pp. 1088-1013, 2004.

[7] W. Wang, X. Zhang, S. Gombault, and S. J. Knapskog, “Attribute Normalization in Network Intrusion Detection,” The 10th International Symposium on Pervasive System, Algorithms, and Networks (ISPAN), 2009, pp. 448-453.

[8] P.Garca-Teodoro, J. Das-Verdejo, G. Maci-fernadez, and E. Vzquez, “Anomaly-based Network Intrusion Detection: Techniques, Systems and Challenges,” Computers and Security, Vol. 28, pp. 18-28, 2009.

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

[10] K. Lee, J. Kim, K.H. Kwon, Y.Han, and S.Kim, “DDoS Attack Detection Method Using Cluster Analysis,” Expert System with Applications, Vol. 34,, no. 3, pp. 1659-1665, 2008.