Multivariate correlation analysis for network intrusion detection

Sumedh P Hanmante, Uma Deshattiwar, Renuka Pawar
Sardar Patel Institute of Technology,Mumbai

Abstract- Interconnected systems, such as Web servers, database servers, cloud computing servers and so on,are now under threats from network attackers. As one of most common and aggressive means, denial-of-service (DoS) attacks cause serious impact on these computing systems. Denial of service (DoS) attack is potential damaging attack which degrades the performance of online servers within seconds. This attack imposes intensive computation on the target server by flooding it with large useless packets. Examples of DoS attacks are Buffer Overflow attaack,Teardrop attack,etc.The target server can be forced out of service from a few minutes to even several days. This causes work down of crucial business services running on the target victim. Also Distributed Denial of Service(DDoS) is also cause security threats for the users.In this attack,attacker uses

multiple comprised systems to attack on a single target. causing denial of service.Examples are-TCP SYN Flood attack,IP Address Spoofing,Smurf attack,etc.Since,DDoS attack is launched from multiple sources,it is often more difficult to detect and block as compared to DoS attack.To cope with such damaging attacks becomes challenge for the researchers. Solution for this attack mainly focuses on the development of network-based detection mechanisms.

In this paper, we present a DoS attack detection system that uses multivariate correlation analysis (MCA) for accurate network traffic characterization by extracting the geometrical correlations between network traffic features.

Index Terms- Distributed denial of service, Denial of service,KDD cup99, triangle area,Multivariate correlation.

IJSER

I. INTRODUCTION

Denial of Service (DoS) attack is one of the most common attacks which causes the serious impact in computing system[1].DoS attacks are class of attacks on targets, which aims at exhausting target resources, thereby denying service to valid users. Denial of service \attack is mainly done in categorize to block a node from receiving genuine data or to block the node entirely from another genuine node. This attack is an attempt to make a machine or network resource unavailable to its intended users by either injecting a computer virus or flooding the network with useless traffic. Computer attack and network attack are the two types of dos attack .To break the serversecurity hackers use DoS attack softening technique. The main targets of DoS attack are web server, application server, database server and communication link. It has become a major threat for current computer networks. Dos attack causes serious damages in services of network, so it is essential to develop a dos attack detection system to protect the services of network. There are two types of network based detection systems, viz. misuse based detection system and anomaly based detection system.

In misuse based detection system attacks are detected by monitoring network activities and looking for matches with the existing attack signatures[2].

In misuse based detection system the database should be kept updated is a laborious task as it is a manual process. So, to overcome these drawbacks of misuse based detection system, anomaly based detection system is developed which is a novelty-tolerant detection system[3]. The manual attack analysis and the frequent update of the attack signature database are avoided in the case of misuse-based detection. In this paper, Our proposed system is for protecting services of network against DoS attacks.. This detection system can provide an effective protection to interconnected systems like web servers, database servers, cloud computing servers etc. by considering their commonality. This system is anomaly based detection system and it employs principles of multivariate correlation analysis (MCA). DoS attack detection system detects known and unknown attacks respectively. To enhance and speed up the process of MCA, triangle area technique is introduced to generate better discriminative features. In this system we are using normalization technique. KDD cup 99 dataset is used for evaluation of DoS attack detection system [4].

II. LITERATURE SURVEY

A. Bro: A System for Detecting Network Intruders in Real-Time We describe Bro, a stand-alone system for detecting network intruders in real-time by passively monitoring a network link over which the intruder's traffic transits[5]. We give an overview of the system's design, which emphasizes high-speed (FDDI-rate) monitoring, real-time notification, clear separation between mechanism and policy, and extensibility. To achieve these ends, Bro is divided into an “event engine” that reduces a kernel filtered network traffic stream into a series of higher level events, and a “policy script interpreter” that interprets eventhandlers written in a specialized language used to express a site's security policy. Event handlers can update state information, synthesize new events, record information to disk, and generate real-time notifications via *syslog*. We also discuss a number of attacks that attempt to subvert passive monitoring systems and defenses against these, and give particulars of how Bro analyzes the four applications integrated into it so far: Finger, FTP, Portmapper and Telnet. The system is publicly available in source code form.

B. An Intrusion-detection model

A model of a real-time intrusion-detection expert system capable of detecting break-ins, penetrations, and other forms of computer abuse is described. The model is based on the hypothesis that security violations can be detected by monitoring a system’s audit records for abnormal patterns of system usage. The model includes profiles for representing the behavior of subjects with respect to objects in terms of metrics and statistical models, and rules for acquiring knowledge about this behavior from audit records and for detecting anomalous behavior. The model is independent of any particular system, application environment, system vulnerability, or type of intrusion, thereby providing a framework for a general-purpose intrusion-detection expert system.

C. A Detailed Analysis of the KDD CUP 99 Data Set

During the last decade, anomaly detection has attracted the attention of many researchers to overcome the weakness of signature-based IDSs in detecting novel attacks, and KDD CUP’99 is the mostly widely used data set for the evaluation of these systems. Having conducted a statistical analysis on this data set, we found two important issues which highly affect the performance of evaluated systems, and results in a very poor evaluation of anomaly detection approaches. To solve these issues, we have proposed a new data set, NSL-KDD, which consists of selected records of the complete KDD data set and does not suffer from any of mentioned shortcomings[6].

D. Attribute Normalization in Network Intrusion Detection

Anomaly intrusion detection is an important issue in computer network security. As a step of data preprocessing, attribute normalization is essential to detection performance. However, many anomaly detection methods do not normalize attributes before training and detection. Few methods considered to normalizing the attributes but the question of which normalization method is more effective still remains. In this paper, we introduce four different schemes of attribute normalization to preprocess the data for anomaly intrusion detection. Three methods, *k*-NN, PCA as well as SVM, are then employed on the normalized data as well as on the original data for comparison of the detection results. KDD Cup 1999 data as well as a real data set collected in our department are used to evaluate the normalization schemes and the detection methods. The systematical evaluation results show that the process of attribute normalization improves a lot the detection performance. The statistical normalization scheme is the best choice if the data set is large. The merits and demerits of the detection methods *k*-NN, PCA and SVM are also analyzed and discussed in this paper to suggest their suitable detection environments.

III. FRAMEWORK

The detection mechanism has three steps-
Step 1- Featuregeneration for each record
Step 2- Multivariate correlation analysis
Step 3- Decision Making

In first step generation of basic feature takes place for an individual record from ingress network traffic. To reduce the overhead of the detection, the destination network is monitored and analyzed, which help detector to protect targeted network.

In second step, multivariate correlation analysis is implemented in which includes "Triangle Area Map (TAM) generation" and "Feature Normalization". To extract the correlation between two distinct features within each traffic record coming from the first step, we applied the Triangle area Map generation technique[7]. In feature normalization module, traffic records get normalized which is given as input to the TAM[8]. To replace the original basic features, all the triangle area correlations stored in triangle area maps (TAMs) are then used. This helps to differentiate between legitimate traffic records. In decision making step, anomaly base detection mechanism is adopted because of which we can detect the attack without requiring any attack relevant knowledge.

In third step called "Decision Making" includes two phases namely training phase and test phase. In training phase "Normal profile generation" is operated in which generation of profiles takes place for various type of legitimate records and are stored in database. In test phase, "Tested profile generation" module is used to build profiles for individual traffic records then tested profile is given to the module called "Attack Detection". The comparison of individual tested profiles with the respective stored normal profiles takes place in the attack detection module.

IJSER

IV. ARCHITECTURE

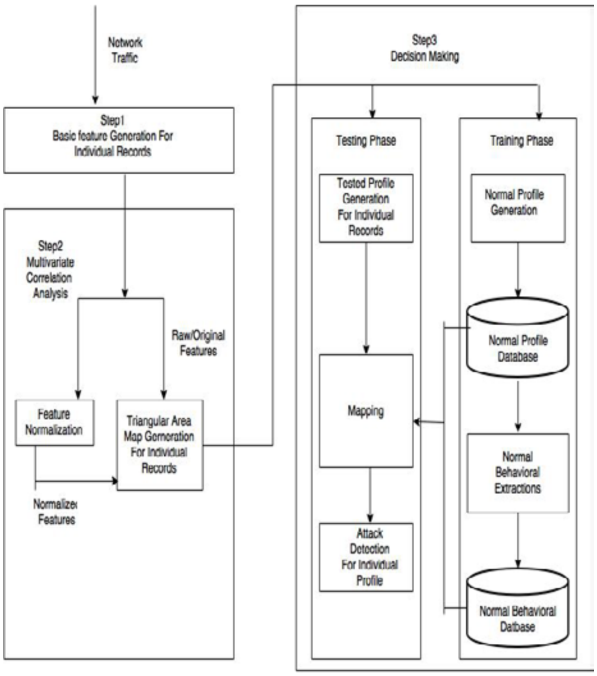
Network Traffic
Step1 Basic feature Generation For Individual Records
Step2 Multivariate Correlation Analysis
Raw/Original Features
Feature Normalization
Triangular Area Map Gemeration For Individual Records
Normalized Features
Step3 Decision Making
Testing Phase
Tested Profile Generation For Individual Records
Mapping
Attack Detection For Individual Profile
Training Phase
Normal Profile Generation
Normal Profile Database
Normal Behavioral Extractions
Normal Behavioral Datbase

Figure 1: Architecture

IJSER

V. MULTIVARIATE CORRELATION ANALYSIS

The behavior of traffic that is attacked by DoS is different from the legitimate network traffic . And the behavior of any network traffic is reflected by its statistical properties. To understand these statistical properties of network traffics, A Multivariate Correlation Analysis approach is there.This MCA approach uses triangle area for extracting the correlative information between the features within an observed traffic records[9]. MCA approach supplies with the following benefits to data analysis. It does not require the knowledge of historic traffic in performing analysis. Unlike the Covariance matrix approaches proposed in, which is vulnerable to linear change of all features, this triangle-areabased MCA withstands the problem. It provides characterization for individual network traffic records rather than model network traffic behavior of a group of network traffic records. This results lower latency in decision making and enable sample-by-sample detection. 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.

A. Detection by anomaly based detection technique-

This section represents a threshold-based anomaly detection method, whose normal profiles are generated using purely legitimate network traffic records . And then can be used for future comparisons with new incoming traffic records. The changes between a new incoming traffic record and the normal profile are examined by the detector .The traffic record is flagged as an attack, If the dissimilarity is greater than a specified threshold. Otherwise, it labeled traffic as a legitimate traffic. Normal profiles and thresholds have direct impact on the performance of a threshold-based detector. A low quality normal profile may cause an inaccurate characterizationto network traffic of in which services are servicing by the servers. Therefore, first apply the proposed triangle area- based MCA approach to analyze legitimate network traffic, and the generated Triangular area map generations are then used to supply features for normal profile generation.

B. Detection by triangle area based technique-

A triangle-area-based MCA approach is applied to analyze the records. Mahalanobis Distance is under concern to measure the dissimilarity between traffic records because MD has been widely used in cluster analysis, classification and multivariate detection techniques. It calculates distance between two multivariate data objects by taking the correlations between variables and removing the dependency.

C.Detection by threshold based technique-

Threshold is specially used to differentiate attack traffic from the legitimate one. Threshold = $\mu + \sigma * \alpha$ W here α denotes normal distribution and usually ranged from 1 to 3. Detection decision can be made with a certain level of confidence, varying from 68% to 99.7% by the selection of different values of α. Finally, if the MD between an observed traffic record and the respective normal profile is greater than the threshold, this will be considered as an attack.

VI. PROPOSED ALGORITHM

```
Require: Observed traffic records X_observed, normal profile_pro:N($\mu,\sigma^2$),TAM_normallower,Cov and α
1: Generate TAM_observedlower for the observed traffic records X_observed
2: MD_observed □ MD(TAM_observedlower , TAM_normallower)
3: If ( $\mu - \sigma * \alpha$) ≤ MD_observed ≤ ($\mu + \sigma * \alpha$), then
MCA Based attack = False
Else
MCA Based attack = True
End if
4: If (MCA Based attack==True)
Return Attack
Else
Generate beh_attack containing behaviors of the attack profile features.
Generate beh_observed containing behaviors of the observed profile features
If beh_attack ≈ beh_observed
return Attack
Else
```

```
Return Normal
End If
```

VII. SOME CALCULATIONS

A. Normal profile Generation

The triangle area based MCA approach is applied to analyze the record. Assume that there is a set of g the training records are

X normal={x1normal,x2 normal, ...x g normal}

B.Mahalanobis Distance(MD)

Measuring the distance between a point P and distribution D. it is a multi dimensional generalization of the idea of measuring many standardvariation away P is from the mean D. This is zero if P is at the mean of D, and grows as p moves away from the mean

MD=sq.root of{(x-μ)tS}-(x- μ)

C.Threshold selection

It is used to separate attack traffic from the legitimate one

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

D.Attack Detection

To detect Dos attacks, the lower triangle of TAM of an observed record needs to be generated using the future triangle- area-based MCA approach.

IJSER

VIII. CONCLUSION

Here we conclude that, the MCA-based DoS attack detection system which uses the triangle area based MCA technique and the anomaly based detection technique is very useful to extracts the geometrical correlations in each individual pairs of two distinct features within each network traffic record and also can be used to extract the correlation between groups of network traffic records. And it gives more accurate characterization for network traffic behaviors. And this technique facilitates computing systems to be able to detect both known and unknown DoS attacks from network traffic.

IX. REFERENCES

[1].https://www.incapsula.com/ddos/ddos-attacks/denial-of-service.html

[2].https://en.wikipedia.org/wiki/Misuse_detection

[3].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

[4].KDD Cup 1999. Available on: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html, Ocotber 2007

[5].Zhiyuan Tan, Aruna Jamdagni, Xiangjian He,Priyadarsi Nanda, Ren Ping Liu , „A System for Denial-of-Service Attack Detection Based on Multivariate Correlation Analysis‟, IEEE Transactions On Parallel And Distributed Systems, Vol. 25, No. 2, February 2014

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

[7]. 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 Int’l Conf.Trust, Security and Privacy in Computing and Comm., pp. 33-40, 2012.

[8].http://agiledata.org/essays/dataNormalization.html

[9].http://www.unb.ca/research/iscx/dataset/iscx-NSL-KDD-dataset.html

AUTHORS

First Author – Sumedh P Hanmantem BEIT, Sardar Patel Institute of Tchnology,Mumbai
Email-hsumedh@gmail.com

Second Author – Uma Deshattiwar, BEIT, Sardar Patel Institute of Technology.,Mumbai
EMAil-usdeshattiwar@gmail.com

Correspondence Author – Renuka Pawar, Assistant Professor, Sardar Patel Institute of Technology,Mumbai