A WEB APPLICATION DETECTING DOS ATTACK USING MCA AND TAM

Pratik Sawant[1], Minal Sable[2], Pooja Kore[3], Shital Bhosale[4]

[1]*BE Student, JSPM's Imperial College Of Engineering And Research, Pune,, India*
[2]*BE Student, JSPM's Imperial College Of Engineering And Research, Pune,, India*
[3]*BE Student, JSPM's Imperial College Of Engineering And Research, Pune,, India*
[4]*BE Student, JSPM's Imperial College Of Engineering And Research, Pune,, India*

Abstract

Interconnected systems, such as all kind of servers including web servers, are been always under the threats of network attackers. There are many popular attacks like man in middle attack, cross site scripting, spamming etc. but Denial of service attack is considered to be one of most dangerous attack on the networked applications. The attack causes many serious issues on these computing systems A denial-of-service (DoS) attack is an attempt to make a machine or network resource unavailable to the intended users. The performance of the server is reduced by the DoS attack, so, to increase the efficiency of the server, detection of the attack is necessary. Hence Multivariate Correlation Analysis' issued, this approach employs triangle area for extracting the correlation information between network traffic. Our implemented system is evaluated using KDD Cup 99 data set, and the treatment of both non-normalized data and normalized data on the performance of the proposed detection system are examined. The implemented system has capability of learning new patterns of legitimate network traffic hence it detect both known and unknown types of DoS attacks and we can say that It is working on the principle of anomaly based attack detection. Triangle-area-based technique is used to speed up the process. The stored legitimate profiles has to keep secured so Detection e=mechanism for the SQL injection is also implemented in the system. The system designed to carry out attack detection is a question-answer portal i.e. a web application and hence the system is using HTTP protocol unlike previous systems which were using TCP.

Keywords: *Denial-of-Service attack, Features Normalization, Triangle Area Map(**TAM**), Multivariate Correlation Analysis(**MCA**), anomaly based detection, **SQL** injection, **HTTP**, and **TCP**,*

--***--

1. INTRODUCTION

Denial-of-service attack is an attempt of increasing the requests traffic and hold all the available resources to make them unavailable for its intended users. This attack consists of efforts to temporarily or indefinitely interrupt or suspend services of a host connected to the Internet. Denial-of-service threats are also common in business, and are sometimes responsible for website attacks. One common method of attack involves saturating the target machine with external communications requests, so much so that it cannot respond to legitimate traffic, or responds so slowly as to be rendered essentially unavailable. Such attacks usually lead to a server overload. The machine or network can be forced out of service even for several days. This can cause damages to the services running on that machine. Therefore it is necessary to detect DOS attacks to protect the online services.

Work on DOS attack detection mainly focuses on the development of network based detection mechanism .This network security mechanism uses either of two main systems viz. misuse-based detection system and anomaly-based detection system. The first one, misuse-based employs signature detection to discriminate between anomaly or attack patterns (signatures) and known intrusion detection signatures .To keep signature database updated is a complicated and labor intensive task ,because signature generation is a manual process and heavily involves network security expertise, therefore research community started to explore a way to achieve novelty-tolerant detection system and developed an advanced concept, namely anomaly based detection system which explores issues in intrusion detection associated with deviations from normal system or user behavior. It is not constrained by the expertise in network security , because the profiles of legitimate behaviors are developed based on techniques, such as data mining and statistical analysis.

We have implemented MCA-based detection system in this paper to protect online services against DoS attacks .For our proposed dos attack detection system developed an algorithm for normal profile generation and an algorithm for attack detection. We proceed a detailed and complete mathematical analysis of the proposed system. Our proposed detection system can provide effective protection to all systems by considering their commonality

2. SYSTEM ARCHITECTURE

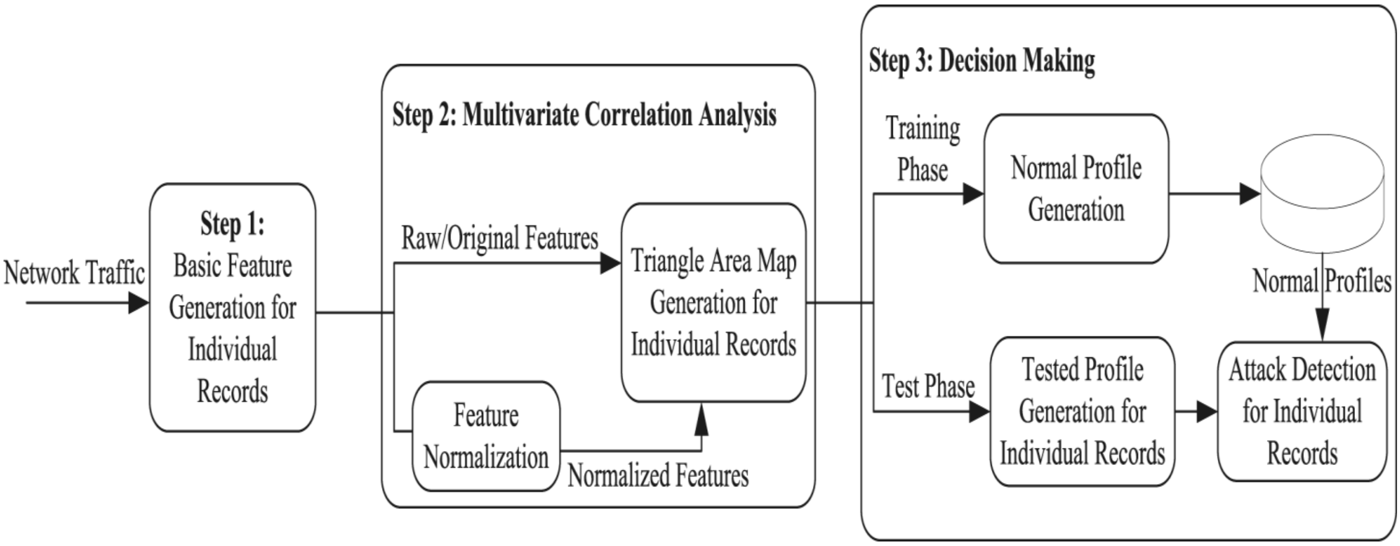
Network Traffic
Step 1:
Basic Feature Generation for Individual Records
Step 2: Multivariate Correlation Analysis
Raw/Original Features
Triangle Area Map Generation for Individual Records
Feature Normalization
Normalized Features
Step 3: Decision Making
Training Phase
Normal Profile Generation
Normal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

Fig. 2.1. System Architecture

The architecture of the system to detect Denial of Service Attack with the modules implemented in the project are discussed in this section.

2.1 Framework

There are three major steps involved in the framework as shown in the architecture diagram viz. Features Generation[1], Multivariate Correlation Analysis[1] and decision making.

Features Generation is the task where the incoming data is divided into smaller parts. For analysis purpose, all the required features are been generated in this phase.

Dividing data in smaller part increases the accuracy of the system in generation of the features. Features Generation is the module necessary for us to implement Sample By Sample detection.

Multivariate Correlation Analysis is the analysis technique used for analysis of the incoming information. It is subdivided into two phases-

1. Features Normalization
2. Triangle Area Map generation

Features Normalization is the process in which we are normalizing the generated features i.e. checking the generated features are matching the previously defined normalization levels. Three normalization levels are defined by us for the normalization purpose which is discussed later.

Triangle Area is the area inside the triangle which is defined using the mathematical module further in the paper. The data inside the triangle is not acceptable by the system where as the data outside the system is acceptable.

Next step to be performed is Decision making. This phase is responsible to state whether the incoming data is an attack or a legitimate request. Decision making is also divided in two phases-

1. Training Phase
2. Test Phase

Training phase is the one where system is made ready by feeding various types of attacks and storing their profiles into database i.e. in triangle area. So that, whenever some new data comes, it will check with the database created in the test phase and make the decision whether the incoming data is an attack or not.

2.2 Sample By Sample Detection

The data generated by the Features Normalization phase i.e. spitted and well structured data is the necessity of the Sample By Sample Detection phase. Since this phase is working on the small amount of data at a time, we achieve following benefits –

a. Accuracy increases as well as the detection is prompt i.e. at early level of attack.
b. New patterns Can be labeled and stored in the triangle map
c. We can classify the sample into its family correctly and hence detection becomes effortless.

3. MCA ANALYSIS

Behavior of the legitimate data is different than the attack traffic. This Behavior is reflected in the various properties of the data. These properties are being examined in the MCA analysis technique.

Consider the data approaching to the system as X = {x1,x2,x3……xn} is the m-dimensional traffic of data. This dataset is divided using normal profile generation by separating x1, x2, x3, … , xn as the separate elements to be treated individually by the sample by sample detection technique.

First of all the normal profile is generated for each of the separated element say Fi = {f1, f2, f3, … fn} where Fi is the normal profile whereas f1, f2… are the number of parameters to be examined during the analysis. Then the triangle data is extracted and compared with the generated profile. Triangle area has different kind of deviated behavior profiles are been stored according to the application and those are the labels according to each sample. Hence the generated profile and the existing profile can be compared to label it. Then the compared profile is forwarded to the Decision Making phase.

As stated above, the various properties are being examined and the previous knowledge of the attacks is not needed in the detection process. When the property doesn't match, it puts an alert and the attack gets detected. Hence the system built comes under the family of Anomaly Based Attack Detection System.

Algorithm 1: Multivariate Correlation Analysis

Here we check the question posted by normal-user is in normalized format and out of Triangle Area Map.

1) Feature Normalization

Step1:NormalizationStepOne

Here we normalize the data (question) according to

i) Question Length i.e.
checkQuestionLength(question)
Which return the length of question.

ii) Word Count(ie. To complete a question we require minimum three words) i.e.
checkWordCount(question)
Which return the word count.

iii) Consecutive words must not be same i.e.
checkSameWords(question)
Which returns Boolean value true or false.

Step 2: Normalization Step Two

Here we normalize the data (question) according to

i) Punctuation Marks (question must not contain only punctuation marks)
i.e. checkPunctuation(question)
which returns Boolean value true or false.

ii) Complete Questiion(A question that starts with wh or do must end with ?)
i.e. checkQuestionMark(question)
which returns Boolean value true or false.

Step 3:NormalizationStepThree

Here we normalize the data (question) according to

i) Only numbers (A question must not contain only numbers)
i.e. checkOnlyNumbers(question)
which return Boolean value true or false.

ii) Only space or tabs (A question must not be blank)
i.e. checkOnlySpacesorTabs(question)
which returns Boolean value true or false.

2) Triangle Area Mapping

Here we map the question posted by normal-user to the triangle are map generated in Algorithm 1

i.e. mapQuestion(question)

which returns Boolean value true or false.

4. DETECTION MECHANISM

Detection mechanism uses the threshold value calculated using the purely legitimate traffic records. Normal profiles are used for this purpose and then are compared with the profiles in the test phase. The dissimilarity between these two profiles is checked. If it is greater than the threshold, then the traffic record is said to be an attack. Since the normal profiles and the threshold is used to flag the traffic as attack, the accuracy of normal profiles is very much important. Otherwise it will result in increasing the false detection rates. Hence the MCA and triangle area map is used to accurately build the normal featured profiles.

4.1 Normal Profile Generation:

In the training phase of the application, normal profiles are been generated using the incoming data. Profile generation is consists of extracting the features of the data/traffic. These profiles are being generated by constructing the lower triangle area map using this data. The dependency is one of the main important factor while calculating the normal profile. Every data has some dependency when it comes in the group. So, by calculating the dependencies, for the normal profiles, by separating the data and checking their dependencies, the independent data is treated separately. Because there is higher possibility in getting the attacking behavior in independent data.

Algorithm 1: Basic Feature Generation for Individual Records

With Administrator of the system having access to questions posted by normal-user, he can generate the Triangle Area Map (TAM) for the system which consists of irrelevant questions.

i.e TAM(Tr1,Tr2…….Tn);

where Tr1 = (f1,f2,f3……fn)

and f1 = {x1,x2….xn} represents the dataset of questions.

4.2 Threshold Selection

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

"α" can vary from 1 to 3 depending on the level of normalization. Threshold selection is very important stage since it directly depends on the false positive rates of attack detection.

4.3 Attack detection

Attack is detected when the TAM of the observed traffic record is compared with the TAM of the normal profile and the comparison have generated value more than the threshold.

Algorithm : Decision Making

Step 1: Training Phase

Here we check the data with the normalized data that we created in Algorithm2

If data doesn't match then normal profiles are generated and question is passed to expert-user for answering.

i.e. insertProfile(question)

which will insert the profile.

Step 2: Test Phase

Here we detect the attack by comparing the data with dataset and if match is found then we deny the service for the user for individual records.

i.e. block(user)

5. EVALUATION

We have designed a web based application that is .question-answer portal for detection of DoS attack . Users are having authentication and security to access the detail which is presented in the our system. Before accessing or searching the details user should have the account in that otherwise they should register first. In this module, there are three users viz. administrator, Expert user, Normal user. Administrator has authority to monitor activities of expert and normal user. Administrator has authority to activate the expert user .Administrator has rights to block or unblock all the normal user who intend to attack the system. Normal users can post the questions and expert users can answer for those questions.

6. CONCLUSION

The system implemented here is useful and efficient technique for the detection of DoS attack. We have extracted important features from MCA (analysis technique) and speeded it up using triangle area map method. The application also has implemented the detection of SQL injection attack to preserve the data needed for the analysis purpose. The future work will be consist of packing methods implemented in the application as the services which will run on the server and prevent any website from the attacks listed above

ACKNOWLEDGEMENTS

We would like to sincerely thank Mrs. Darshika Lothe, our guide, for her support and guidance. We are also thankful of our HOD Prof. S. R. Todmal and Principal Dr. S.V. Admane for his support and encouragement. The precious guidance is also provided by our other faculty members time to time hence, we are also thankful to them

REFERENCES

[1] http://en.wikipedia.org/wiki/Denial-of-service_attack

[2] www.computer.org/csdl/trans/td/2014/02/ttd2014020447-abs.html

[3] http://www.cse.litm.ac.in/~ravi/paper/vijaysarathy_thesis.pdf

[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, 2008.

[5] A. Tajbakhsh, M. Rahmati, and A. Mirzaei, "Intrusion detection using fuzzy association rules," Applied Soft Computing, vol. 9, no. 2, pp. 462-469, 2009.

[6] J. Yu, H. Lee, M.-S. Kim, and D. Park, "Traffic flooding attack detection with SNMP MIB using SVM," Computer Communications, vol. 31, no. 17, pp. 4212-4219, 2008.

[7] 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, 2008.

[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, 2007.