Multivariate Correlation Analysis Technique for DoS Attack Detection

D. Neela [1]
[1]Student, Department of CSE,
JJ College of Engineering and Technology,
Tamilnadu, India.

P. D Sheba Kezia Malar Chelvi [2]
[2]Professor, Department of CSE,
JJ College of Engineering and Technology,
Tamilnadu, India.

***Abstract*— In the recent years, Denial of Service (DoS) attacks have been widely spread threats to network security. DoS attack is an attempt to make a machine or network resources unavailable to its intend users. In this paper, the various methods available in the literature for detection of DoS attacks are analyzed and also one of the recent technique to detect DoS attacks based on a statistical approach namely Multivariate Correlation Analysis (MCA) is explored. MCA technique employs triangle area for extracting the geometrical correlation information between the network traffic features. MCA-based DoS attack detection system employs the principle of anomaly-based detection in attack recognition. This makes the solution capable of detecting known and unknown DoS attacks effectively by learning the patterns of legitimate network traffic only. The proposed system will be evaluated using KDD Cup 99 data set.**

Keywords*— *Denial of service attack; multivariate correlation; triangle area; anomaly detection;

I. INTRODUCTION

Denials of Service attacks are a kind of attacks against computers connected to the internet. DoS attacks exploit bugs in a specific operating system or vulnerabilities in TCP/IP implementation. Unlike a privacy attack, where an adversary is trying to get access to resources to which it has no authorization, the goal of DoS attacks is to keep authorized users from accessing resources. The infected computers may crash or disconnect from the internet. In some cases they are not very harmful, because once you restart the crashed computer everything is on track again. In other cases they can be disasters, especially when you run a corporate network or ISP.

According to the detection strategy used, detection systems can be classified into two main categories. They are misuse detection, which identifies intrusions using patterns of well known intrusions or weak spots of the system and anomaly detection, which attempts to find out if departure from the recognized standard usage patterns can be flagged as attacks. Misuse Detection tries to model abnormal activities from impressions of known intrusions and known system weaknesses. In Anomaly Detection Anomaly detectors identify possible attack attempts by constructing profiles representing normal usage and then comparing it with current behavior data to find out a likely mismatch. As in [5] anomaly detection techniques are classified according to the nature of processing and the same is depicted in *Fig.1*.

A. Statistics Based

In statistics-based techniques, the network traffic activity is captured and a profile indicating its stochastic behavior is created. This profile is based on metrics such as the traffic rate, the number of packets for each protocol, the rate of connections, the number of different IP addresses, etc. Two datasets of network traffic are well thought-out during the anomaly detection process: one corresponds to the currently observed profile over time, and the other is for the previously trained statistical profile. . As the network events occur, the current profile is determined and an anomaly score estimated by comparison of the two behaviors. The score usually indicates the degree of abnormality for a specific event, such that the intrusion detection system will flag the occurrence of an anomaly when the score surpasses a certain threshold.

B. Knowledge Based

The so-called expert system approach is one of the most widely used knowledge-based IDS schemes. Expert systems are intended to classify the audit data according to a set of rules, involving three steps. First, different attributes and classes are identified from the training data. Second, a set of classification rules, parameters or procedures are deduced. Third, the audit data are classified accordingly. More restrictive/particular in some senses are specification-based anomaly methods, for which the desired model is manually constructed by a human expert, in terms of a set of rules (the specifications) that seek to determine legitimate system behavior. If the specifications are complete enough, the model will be able to detect illegitimate behavioral patterns. Moreover, the number of false positives is reduced, mainly because this kind of system avoids the problem of harmless activities, not previously observed, being reported as intrusions.

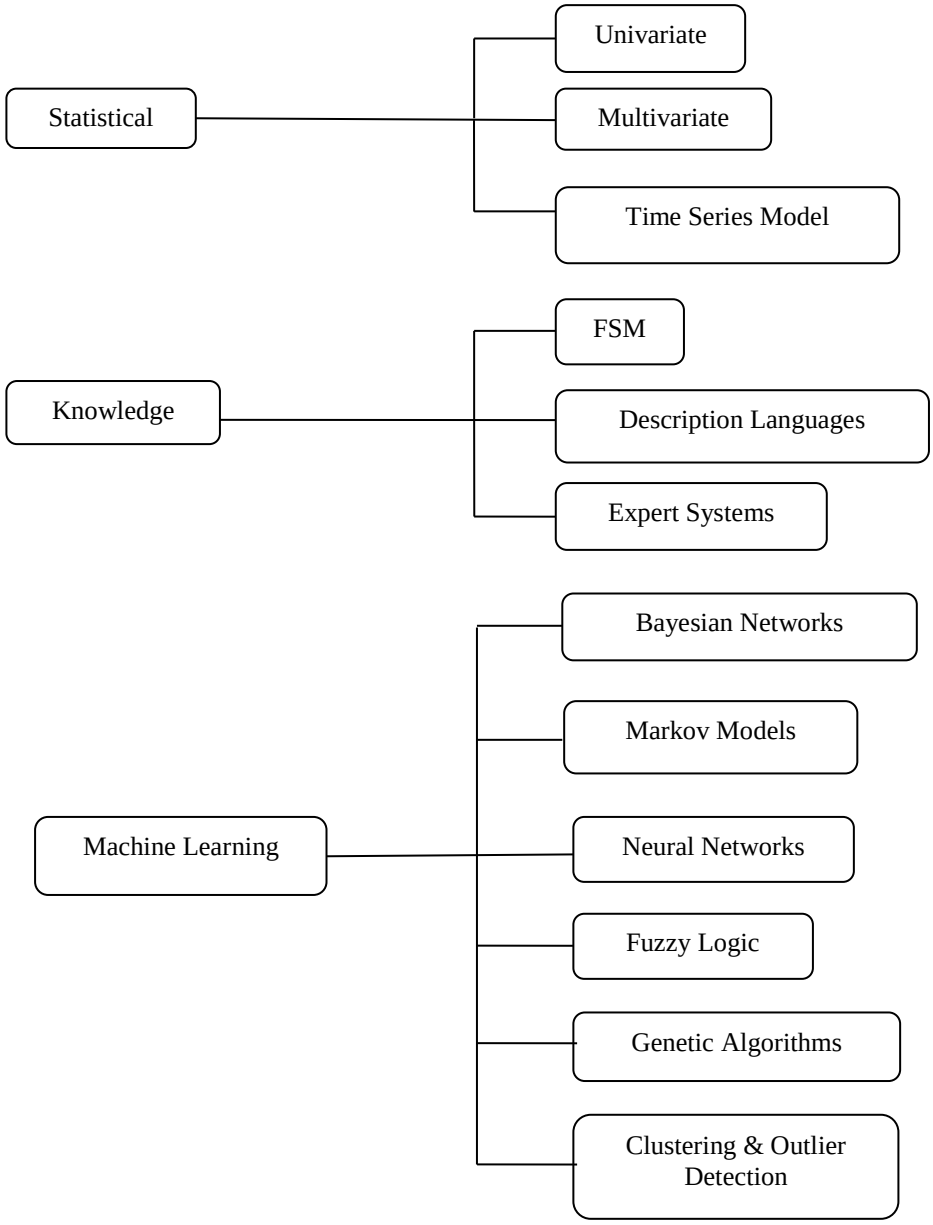

Figure:1 Classification of the anomaly detection techniques according to the nature of the processing involved in the ‘‘behavioural’’ model considered

C. Machine Learning Based

Machine learning techniques are based on establishing an explicit or implicit model that enables the patterns analyzed to be categorized. A singular characteristic of these schemes is the need for labelled data to train the behavioural model, a procedure that places severe demands on resources. In many cases, the applicability of machine learning principles coincides with that for the statistical techniques, although the former is focused on building a model that improves its performance on the basis of previous results. Hence, a machine learning A-NIDS has the ability to change its execution strategy as it acquires new information. Although this failure could make it desirable to use such schemes for all situations, it is very expensive.

The remainder of this paper are as follows: Section 2 presents a Related work for different types of DoS attacks and its detection techniques. Such as i) Bandwidth utilization pattern and rule based pattern matching ii) Detection against Domain Name System using Machine Learning Classifiers iii) Network Intrusion Detection based on Misuse Detection iv) Detection using Principal component analysis v) detection using Distributed Denial of service Detection Mechanism (DiDDeM). Section 3 presents the proposed system and finally conclusion is given in section 4.

II. RELATED WORK

Challa Madhavi et al. (2011) have proposed a technique to detect the intrusion based on bandwidth usage pattern analysis combined with protocol headers pattern matching of the packets that are being exchanged from the system with the internet or network. The system comprises of mainly three components: a monitor which senses and extracts the packet information from the packets being exchanged, classifier

classifies the packets as being intruding and non intruding and performance analyzer to analyze the system.

Iqbal Saripan M. et al. (2010) largely focused on the detection method based on machine learning techniques. Domain Name System (DNS) provides name to address mapping services for the entire chain of internet connectivity. Hackers exploit this fact to damage different parts of the Internet. The system consists of a statistical pre-processor and a machine learning (ML) engine. Three different types of neural network classifiers namely BP neural network, RBF neural network, SOM neural network and support vector machines are utilized. Optimized BP network that can effectively detect and classify different DoS attacks against DNS. To implement an optimized RBF neural network for classification problem, specify the activation function for the hidden units have been specified and the centre and widths of RBFs. In SOM neural network, the input vector of three features has been normalized due to the large variations of input values.

Anuradha et al. (2011) explained about the Intrusion Detection System (IDS) is the process of monitoring the events occurring in a computer system or network and analyzing them for signs of possible incidents, which are violations or imminent threats of violation of computer security policies, acceptable use policies, or standard security practices. Incidents have many causes such as malware (e.g., worms, spyware), attackers gaining unauthorized access to systems from the Internet, an authorized users of systems who misuse their privileges or attempt to gain additional privileges for which they are not authorized. It uses signature for attack detection. The system not only detects the intruders by the IP address, it detects the system with its contents also. The system checks the database for the already registered intruders. If found intruding, they are forwarded to the firewall for blocking.

Khaled Labib et al. (2010) have proposed a multivariate statistical method called Principal Component Analysis to detect selected Denial-of-Service and network probe attacks. The principal components are calculated for both attack and normal traffic, and the loading values of the various feature vector components are analyzed with respect to the Principal Components. The variance and standard deviation of the Principal Components are calculated and analyzed.

Prathap A. et al. (2012) have proposed an early detection system that can detect the presence of a worm in the Internet as quickly as possible in order to give people accurate early warning information and possible reaction time for counteractions. Distributed DoS Detection Mechanism (DiDDeM) providing the means by which DoS attacks are detected early. The three key elements of DiDDeM are: the DiDDeM domain, the pre-filter (PF) detection node, and the command and control (C2) server. A DiDDeM domain comprises a number of PF nodes and a C2 server. A PF is a key element in the detection of denial-of-service attacks. A PF is located on a router and utilizes the congestion algorithm to infer stateful information from stateless information for detection. A C2 server manages its DiDDeM domain, communicates with C2 servers in other DiDDeM domains, and provides a central station for attack analysis, and coordinates a response to an attack.

III. PROPOSED SYSTEM

The work proposed in this paper utilizes the triangle area map generation technique proposed by Aruna Jamdagni et al. (2012) for performing correlation analysis between the normal profile and the input records to identify attacks.

A. Feature Attribute Selection

The classification of the 41 features of the dataset sorted in a descending order is made through the information gain ratio as in [6]. Most of the features have Information Gain Ratio (IGR) under the average of the data set, (IGR average = 0.22). In fact, only 20 features are above the average. Selected Attributes for Individual Attack are categorized as below:

Attacks	**Selected Attributes**
DoS	3,4,5,6,8,10,13,23,24,37

B. Feature Normalization

There are 3 features in each packets have characters values (protocol type, Service, Flag), which must be converted to numeric value as described in [7]. For instance, in the case of protocol type feature, 0 is assigned to TCP, 1 to UDP, and 2 to the ICMP symbol at the same way other fields (Service , Flag) are coded. also to scale between [0,1] and the method used for this process is given in [8]. For example ‘protocol_type' it has the following string values { TCP, UDP, UDP, UDP, RTP, RTP, ICMP, TCP, TCP}, then we have "M=9 and K=4" different strings in the feature. Probability mass function is defined for each string such as pmf(TCP) = 3/9 = 0.33, pmf (UDP) = 3/9 = 0.33, pmf (RTP) = 2/9 = 0.22, pmf (ICMP) = 1/9 = 0.11. As a result the nominal feature have been transferred to the following numeric representation pmf(X2)= { 0.33, 0.33, 0.33, 0.33, 0.22, 0.22,0.11, 0.33,0.33}. *nv* to be the normalized feature value after normalization process. Statistical Normalization is defined in (1).

$$nv = \frac{v - \mu}{\sigma} \quad (1)$$

where μ and σ are the mean value and standard deviation of a feature vector respectively.

C. Multivariate Correlation Analysis

“Triangle Area Map Generation” module is applied to extract the correlations between two distinct features within each raw/original traffic record coming from the initial step or the traffic record normalized by the " Feature Normalization " module in this step. The occurrence of network intrusions cause changes to these correlations so that the changes can be used as indicators to identify the intrusive activities. All the extracted correlations, namely triangle areas stored in Triangle Area Maps (TAMs), are then used to replace the original basic features or the normalized features to represent the traffic records. This provides higher discriminative information to

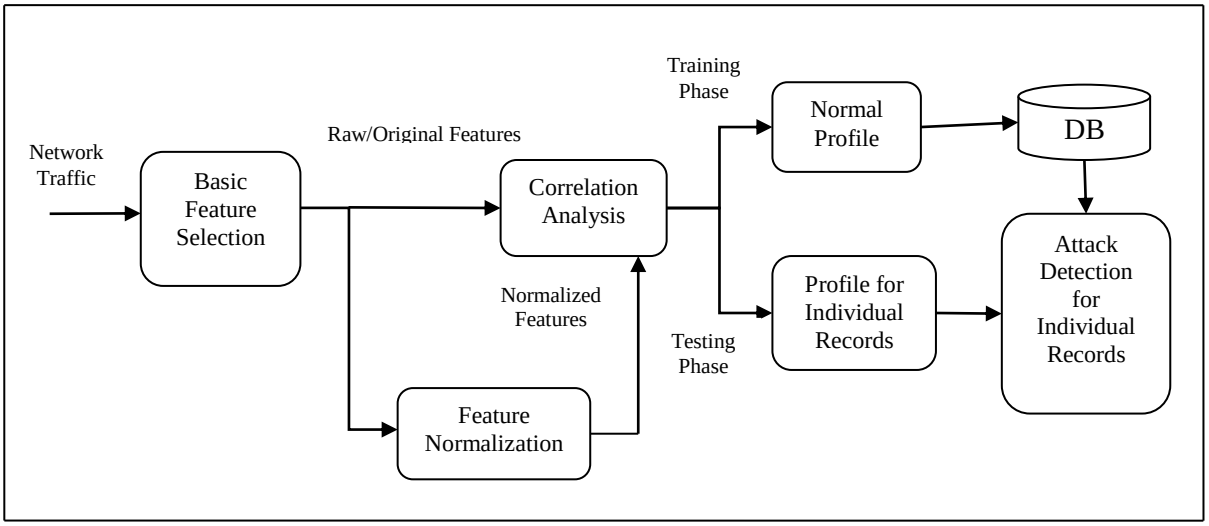

Figure: 2 A System Framework for Denial of Service Attack Detection

differentiate between legitimate and illegitimate traffic records.

D. *Normal Profile Generation*

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. It evaluates distance between two multivariate data objects by taking the correlations between variables into account and removing the dependency on the scale of measurement during the calculation. the distribution of the MDs is described by two parameters, namely the mean μ and the standard deviation σ of the DIST. Finally, the obtained distribution $N(\mu, \sigma 2)$ of the normal training traffic records, MAP_{lower}^{normal} and Cov are stored in normal profile for attack detection.

E. *Tested Profile Generation*

The "Tested Profile Generation" module is used in the "Test Phase" to build profiles for individual observed traffic records. Then, the tested profiles are handed over to the "Attack Detection" module, which compares the individual tested profiles with the respective stored normal profiles. A threshold-based classifiers employed in the "Attack Detection" module to distinguish DoS attacks from legitimate traffic.

F. *Attack Detection*

A threshold-based anomaly detector, whose normal profiles are generated using purely legitimate network traffic records and utilized 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. If the dissimilarity is greater than a pre-determined threshold, the traffic record is flagged as an attack. Otherwise, it is labelled as a legitimate traffic record. Clearly, normal profiles and thresholds have direct influence on the performance of a threshold-based detector.

G. *Algorithm*

Require: Observed traffic record $x^{observed}$, normal profile Pro : $(N(\mu, \sigma 2), MAP_{lower}^{normal}, Cov)$ and parameter α

1:Generate $MAP_{lower}^{observed}$ for the observed traffic record $x^{observed}$

2:$DIST^{observed} \leftarrow DIST(MAP_{lower}^{observed}, MAP_{lower}^{normal})$

3: **if** $(\mu - \sigma * \alpha) \leq DISTobserved \leq (\mu + \sigma * \alpha)$ **then**

4: **return** Normal

5: **else**

6: **return** Attack

7: **end if**

IV. CONCLUSION

In Multivariate correlation based denial of service attack detection system based on triangle technique evaluates network traffic dataset for verification of the effectiveness and performance of proposed system. The former technique extracts the correlations for individual pairs of two distinct features within each network traffic records and characterize more accurate irrelevant behaviors. The further techniques facilitate our system for distinguishing both known and unknown attacks. The original and normalized influence for results reveals working without normalized data with our detection system achieves maximum detection accuracy. This problem solved by utilizing statistical normalization techniques for eliminating bias from data. Our proposed system achieves equal or better performance which compares two state of art approach for denial of service attack. The sophisticated classifiers of techniques for false positive rate which have computational complexity and time taken for detection system.

REFERENCES

[1] Anuradha and Munish Sharma, "Network Intrusion Detection System for Denial of Service Attack based on Misuse Detection" IJCEM International Journal of Computational Engineering & Management, Vol. 12, 2011.

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

[3] Aruna Jamdagni, Priyadarsi Nanda, Ren Ping Liu, Xiangjian He and Zhiyuan Tan, "Triangle-Area-Based Multivariate Correlation Analysis for Effective Denial-of-Service Attack Detection" IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications, 2012.

[4] Challa Madhavi, Dr.Manjunath Gadiparthi, Revathi Cherukuri and Thadoor Shobha Rani, "A Real Time DOS Attack Detection in IP Networks Based on Bandwidth Utilization Pattern and Rule Based Pattern Matching" IJCST Vol. 2, Issue 3, 2011.

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

[6] Firas S. Jassim and Safaa O. Al- mamory, "Evaluation of Different Data Mining Algorithms with KDD CUP 99 Data Set " Journal of Babylon University/Pure and Applied Sciences ,Vol.21, No.8,2013.

[7] Iqbal Saripan M., Mohd Fadlee A. Rasid and Samaneh Rastegari, " Detection of Denial of Service Attacks against Domain Name System Using Machine Learning Classifiers" Proc of the World Congress on Engg.Vol.1, 2010.

[8] Khaled Labib and V. Rao Vemuri, "Detecting And Visualizing Denial-of Service And Network Probe Attacks Using Principal Component Analysis" Department of Applied Science University of California, Davis U.S.A, 2010.

[9] Maher Salem and Ulrich Buehler, "Mining Techniques In Network Security To Enhance Intrusion Detection Systems" International Journal of Network Security & Its Applications (IJNSA), Vol.4, No.6, 2012.

[10] Prathap A. and Sailaja R., "Detection and Prevention of Denial of Service Attacks Using Distributed Denial-of-Service Detection Mechanism" International Journal of Computer Science and Information Technologies, Vol. 3,2012.