IJCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Volume-4, Issue-3 E-ISSN: 2347-2693

Denial of Service Attack Detection Using Multivariate Correlation Information and Support Vector Machine Classification

Subhash Pingale[1], Ranjeetsingh Parihar[2] and Prajakta Solankar[3*]

[1,2,3*] Dept. of Computer Science and engineering

SKN Sinhgad College of Engineering Korti, Pandharpur, Maharashtra

www.ijcseonline.org

Received: Feb/22/2016 Revised: Mar/06/2016 Accepted: Mar/14/2016 Published: Mar/31/2016

Abstract— Denial of service attack (DoS) is serious threat to the internet. The DoS attack affects on the computing systems such as database server, web server etc. Denial of Service attack prevents authorized user from accessing online services. Therefore effective detection of DoS attack is necessary for increasing the efficiency of server. The Multivariate correlation analysis(MCA) for network traffic characterization overcomes the problem of DoS attack. MCA uses triangle area technique for extracting correlative information between network traffic. Triangle area based method is used to speed up the MCA process. Then Support Vector Machine based classification technique used for attack classification using the triangle area based multivariate correlation information. The min-max normalization method presented to increase the detection rate of DoS attack.

Keywords— Multivariate correlation analysis, DoS attack, support vector machine, Triangle area technique, normalization, detection rate

I. Introduction

Denial of Service Attack (DoS) is the serious intrusive behavior for an online server. When DoS attack occurs then authorized user not able to use the online services. The attacker's main target is to disrupt the services provided. Some companies also perform dos attack because of competition in the market. DoS attack also affected on the eBay.com, yahoo.com, amazon.com and also similar other websites. Network Attack Detection Techniques are described below.

A. Misuse based Technique

Misuse based detection technique monitors the network activities and uses the previously stored attack signature for attack detection. To keep the signature database updated is complicated task because we have to manually create the signatures. So generally researchers use anomaly detection mechanism for attack detection.

B. Anomaly based Technique:-

In anomaly based detection technique not necessary to keep the signature database instead it creates the profiles of the legitimate traffic. In this technique profile of legitimate traffic are developed using techniques statistical analysis, machine learning and data mining etc.

Objective: The objective is to provide detection against denial-of-service attack and to improve the performance of detection system.

The fig.1 shows denial of service attack happened on bank. The attacker sends command for his bots to attack a bank. Thousands of requests are sent to bank website simultaneously. Then the bank is flooded with many requests and can't operate effectively.

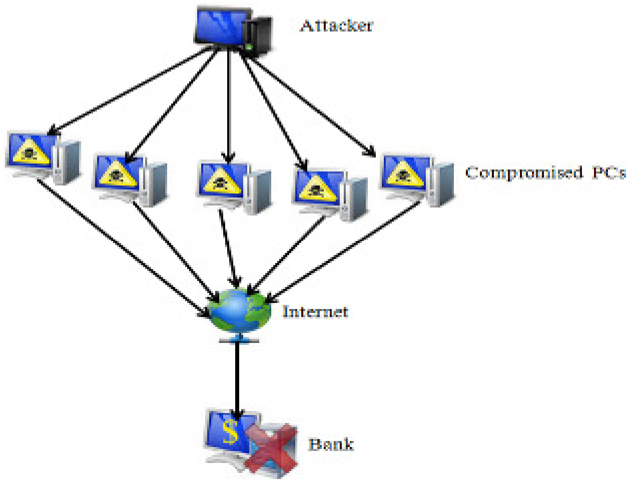


Fig.1 Example of Denial of Service attack on bank

II. Related Work

Aruna Jamdagni et.al. Proposed RePIDS for intrusion detection. RePIDS is the real time payload based detection system. For preprocessing of the data principle component analysis method and MDM (Mahalnobis Distance Map) approach is used to determine the hidden correlation among various features. They used Georgia Institute of Technology attack dataset and DARPA 99 dataset. The traffic of Web

Corresponding Author: *prajakta solankar,prajaktasolankar@gmail.com*
Department of CSE, University of Solapur ,maharashtra., India.

based application is used in this work. They generated Mahalnobis distance map of HTTP payload.

Chih-Fong Tsai and Chia-Ying Lin proposed Triangle area based nearest neighbor approach for detecting intrusions. This work consists of k-means and is used as clustering method to find the center clusters for each category. Author of this paper determined triangle areas. At the end they used k-nearest neighbor classifier for classification [1].The author said that feature selection also necessary to improve the efficiency. They proposed Information Gain and Triangle area based KNN for selecting feature. They used Greedy K-means clustering algorithm and SVM (support vector machine) classifier for detecting network attacks [2].The author proposed algorithm for discriminating DDoS attacks from flash crowds using flow correlation coefficient among suspicious flows. They presented theoretical proofs for the feasibility of the discrimination method in theory. Their aim is to protect victims like web servers, mail Servers and from flash crowd attacks within a community network [3].This work analyzed different methods of attribute normalization for preprocessing the data during anomaly intrusion detection. They described different normalization techniques [4].Jin et al. discussed the effects of MCA for detection of Distributed DoS attacks. For detection of flooding attacks they proposed covariance analysis model. The method separates the traffic as attack or normal. They verified the effect of multivariate correlation analysis (MCA) using covariance model [5]. In the proposed work of Euclidean distance map based multivariate correlation analysis to detect denial of service attack. The Euclidean distance technique used to extract the correlative information from the original feature space of an observed data object [6]. B. Kiranmai and A. Damodaram described various methods for intruder detection [7].

III. FRAMEWORK

Stage1- Firstly kddcup99 dataset network features are collected and then preprocessing on the dataset by using the feature selection and normalization method is performed.

Stage2- After selection of features and normalization, the multivariate correlation of these features using triangle area technique is determined. Triangle areas are used for extracting the correlative information between the features of traffic record and extracted correlations are stored in the Maps and this map is named as TAM i.e. triangle area map. Multivariate correlation analysis is to characterize the network traffic. Triangle area technique is used to speed up the MCA (multivariate correlation analysis) process.

In Triangle area technique if any differences found in the lower triangles can be found in upper triangle also. So use either lower triangle or upper triangles of triangle area map. The MCA method has several advantages, the proposed triangle area based MCA is not vulnerable to linear change of all the features. Geometrical structure analysis reveals the correlations between two distinct features. When anomaly behavior appears in the network then changes occur in the correlations. The results of Triangle area based MCA (multivariate correlation analysis) is given to the last stage i.e. attack detection step for decision making [8]. In MCA covariance matrix, Mahalnobis distance is calculated.

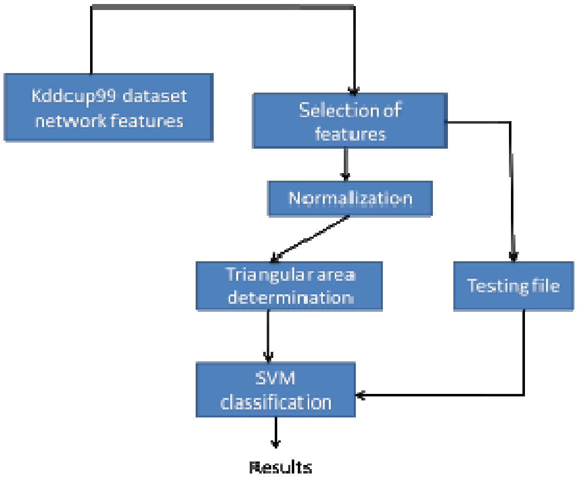


Fig.2. Framework of proposed DoS attack detection system

Stage3- The last stage is to detect attack. In this step multivariate correlation analysis is used by the Support vector machine technique for making decision.SMO (sequential minimal optimization) function of SVM is used for classification. In the attack detection stage testing files area tested by using training set.

IV. MULTIVARIATE CORRELATION ANALYSIS

The traffic of denial of service attack behaves differently from the legitimate network traffic and network traffic behavior is reflected through statistical properties. To well describe statistical properties, multivariate correlation analysis (MCA) information is used. The MCA employs the triangle area method to extract the geometrical correlation among features within traffic record. These correlations are extracted from the statistical properties (eg. the Number of bytes from source to destination and length of connection). In MCA triangle area technique is used to extract the multivariate correlations. The triangle area is determined by using the features of the network and after determining the area, Triangle areas are stored with respect to their indexes (This is TAM). TAM is the symmetric matrix having upper and lower triangular matrix is same. That's why only lower triangular matrix is considered for the profiles generation. Then these triangle area maps are used to develop the normal profiles for attack detection [8].
The area of triangle is calculated by using formula consists of j and k is features:

$$TA^{i}_{j,\,k=\,(|f^{i}_{j}|\,|f^{i}_{k}|)/2.} \quad \ldots\ldots..(1)$$

The triangle areas of $TA^i_{j,k}$ and $TA^i_{k,j}$ is same. Then triangle area map is symmetric matrix with diagonal elements is zeros. All the triangle areas are determined and stored in TAM (triangle area map). When we compare two triangle area maps then we consider only lower triangle of matrix because any differences found on lower triangles is also found on upper triangle. Then lower triangular matrix TAM^i_{lower} = [$TA^i_{2,1}$ $TA^i_{3,1}$... $TA^i_{m,1}$ $TA^i_{3,2}$ $TA^i_{4,2}$...... $TA^i_{m,2}$......... $TA^i_{m,m-1}$] .This matrix is the correlation [8].

V. SUPPORT VECTOR MACHINE

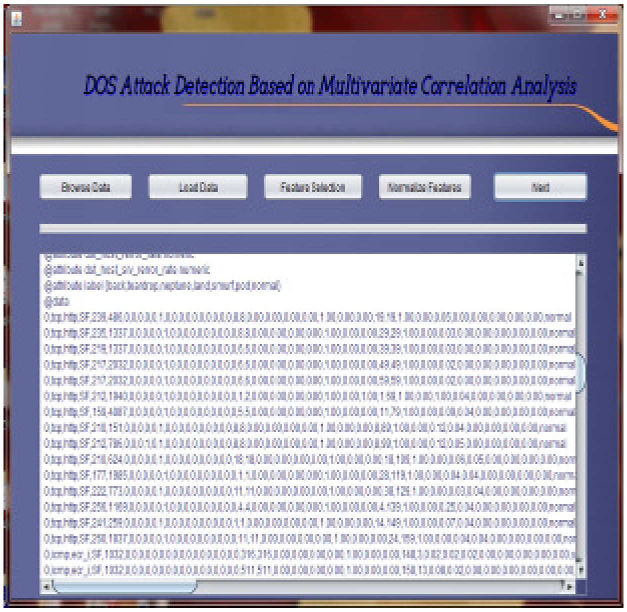
The support vector machine (SVM) is used for the classification of traffic. The support vector machine separates data points into two classes. By using SVM technique Attack or normal traffic is determined. It consists of separating hyperplane which is used to separate data. The points nearest to separating hyperplane are called as support vectors. In this need to maximize the distance between the separating hyperplane and Support Vectors. Support vector machine is first suggested by vapnik for classification. It uses the principle of structural risk minimization [9] [10].

VI. IMPLEMENTATION

In the implementation of the system, Java language is used for development. So here six Modules as defined:

A. Input dataset

Kddcup99 dataset is used as input to the proposed system which consists of different types of records: normal, smurf, teardrop, Neptune etc. There are 42 features in the dataset. The features of dataset are duration, protocol type, service, source address, destination address, wrong fragment, count, urgent etc. The kdd dataset features are shown in below fig.3.

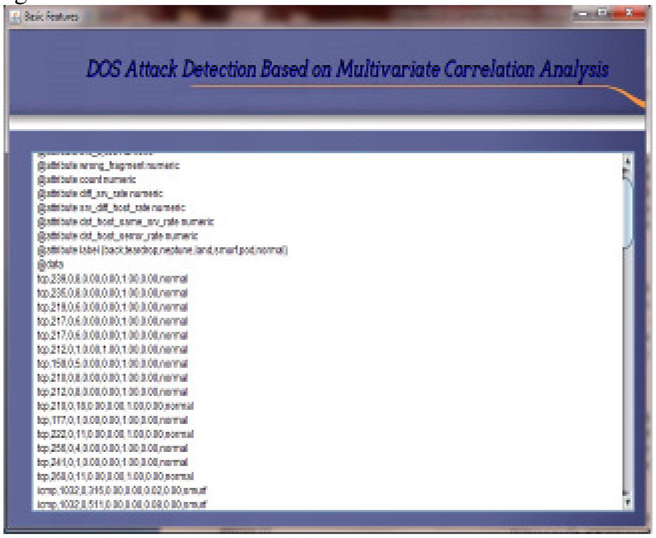

DOS Attack Detection Based on Multivariate Correlation Analysis
Browse Data
Load Data
Feature Selection
Normalize Features
Next

```
@attribute dst_host_srv_serror_rate numeric
@attribute label {back,teardrop,neptune,land,smurf,pod,normal}
@data
0,tcp,http,SF,239,486,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,8,8,0.00,0.00,0.00,0.00,1.00,0.00,0.00,19,19,1.00,0.00,0.05,0.00,0.00,0.00,0.00,0.00,normal
0,tcp,http,SF,235,1337,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,8,8,0.00,0.00,0.00,0.00,1.00,0.00,0.00,29,29,1.00,0.00,0.03,0.00,0.00,0.00,0.00,0.00,normal
0,tcp,http,SF,219,1337,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,6,6,0.00,0.00,0.00,0.00,1.00,0.00,0.00,39,39,1.00,0.00,0.03,0.00,0.00,0.00,0.00,0.00,normal
0,tcp,http,SF,217,2032,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,6,6,0.00,0.00,0.00,0.00,1.00,0.00,0.00,49,49,1.00,0.00,0.02,0.00,0.00,0.00,0.00,0.00,normal
0,tcp,http,SF,217,2032,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,6,6,0.00,0.00,0.00,0.00,1.00,0.00,0.00,59,59,1.00,0.00,0.02,0.00,0.00,0.00,0.00,0.00,normal
0,tcp,http,SF,212,1940,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,2,0.00,0.00,0.00,0.00,1.00,0.00,1.00,1,69,1.00,0.00,1.00,0.04,0.00,0.00,0.00,0.00,normal
0,tcp,http,SF,159,4087,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,5,5,0.00,0.00,0.00,0.00,1.00,0.00,0.00,11,79,1.00,0.00,0.09,0.04,0.00,0.00,0.00,0.00,normal
0,tcp,http,SF,210,151,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,8,8,0.00,0.00,0.00,0.00,1.00,0.00,0.00,8,89,1.00,0.00,0.12,0.04,0.00,0.00,0.00,0.00,normal
0,tcp,http,SF,212,786,0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,8,8,0.00,0.00,0.00,0.00,1.00,0.00,0.00,8,99,1.00,0.00,0.12,0.05,0.00,0.00,0.00,0.00,normal
0,tcp,http,SF,210,624,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,18,18,0.00,0.00,0.00,0.00,1.00,0.00,0.00,18,109,1.00,0.00,0.06,0.05,0.00,0.00,0.00,0.00,norm
0,tcp,http,SF,177,1985,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,1,0.00,0.00,0.00,0.00,1.00,0.00,0.00,28,119,1.00,0.00,0.04,0.04,0.00,0.00,0.00,0.00,norm
0,tcp,http,SF,222,773,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,11,11,0.00,0.00,0.00,0.00,1.00,0.00,0.00,38,129,1.00,0.00,0.03,0.04,0.00,0.00,0.00,0.00,norm
0,tcp,http,SF,256,1169,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,4,4,0.00,0.00,0.00,0.00,1.00,0.00,0.00,4,139,1.00,0.00,0.25,0.04,0.00,0.00,0.00,0.00,normal
0,tcp,http,SF,241,259,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,1,0.00,0.00,0.00,0.00,1.00,0.00,0.00,14,149,1.00,0.00,0.07,0.04,0.00,0.00,0.00,0.00,normal
0,tcp,http,SF,260,1837,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,11,11,0.00,0.00,0.00,0.00,1.00,0.00,0.00,24,159,1.00,0.00,0.04,0.04,0.00,0.00,0.00,0.00,no
0,icmp,ecr_i,SF,1032,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,316,316,0.00,0.00,0.00,0.00,1.00,0.00,0.00,148,3,0.02,0.02,0.02,0.00,0.00,0.00,0.00,0.00,s
0,icmp,ecr_i,SF,1032,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,511,511,0.00,0.00,0.00,0.00,1.00,0.00,0.00,158,13,0.08,0.02,0.08,0.00,0.00,0.00,0.00,0.00,s
```

Fig.3. Kddcup99 dataset instances

B. Feature selection

In feature selection only useful features are selected. we selected relevant features for dos attack classification. Feature selection is the preprocessing of the data. we selected protocol, source address, destination address, wrongfragment,count,diff_srv_rate,srv_diff_host_rate,dst_host_same_srv_rate, dst_host_serror_rate..This is shown in fig.4.

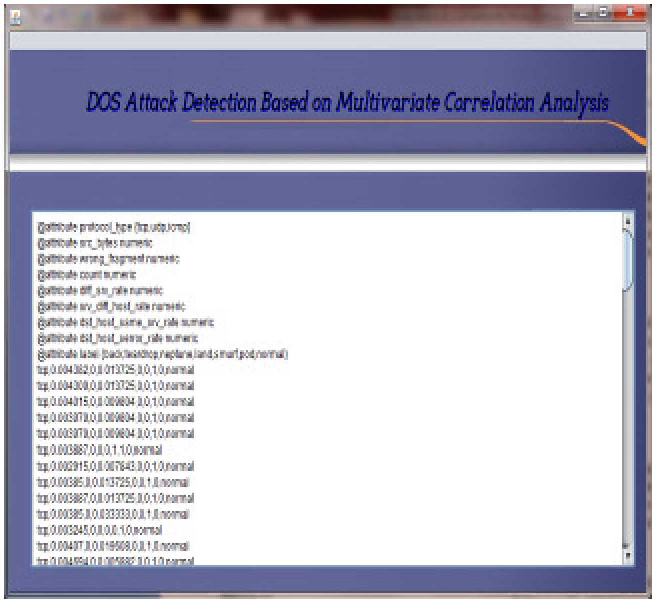

Basic Features
DOS Attack Detection Based on Multivariate Correlation Analysis

```
@attribute wrong_fragment numeric
@attribute count numeric
@attribute diff_srv_rate numeric
@attribute srv_diff_host_rate numeric
@attribute dst_host_same_srv_rate numeric
@attribute dst_host_serror_rate numeric
@attribute label {back,teardrop,neptune,land,smurf,pod,normal}
@data
tcp,239,0,8,0.00,0.00,1.00,0.00,normal
tcp,235,0,8,0.00,0.00,1.00,0.00,normal
tcp,219,0,6,0.00,0.00,1.00,0.00,normal
tcp,217,0,6,0.00,0.00,1.00,0.00,normal
tcp,217,0,6,0.00,0.00,1.00,0.00,normal
tcp,212,0,1,0.00,1.00,1.00,0.00,normal
tcp,159,0,5,0.00,0.00,1.00,0.00,normal
tcp,210,0,8,0.00,0.00,1.00,0.00,normal
tcp,212,0,8,0.00,0.00,1.00,0.00,normal
tcp,210,0,18,0.00,0.00,1.00,0.00,normal
tcp,177,0,1,0.00,0.00,1.00,0.00,normal
tcp,222,0,11,0.00,0.00,1.00,0.00,normal
tcp,256,0,4,0.00,0.00,1.00,0.00,normal
tcp,241,0,1,0.00,0.00,1.00,0.00,normal
tcp,260,0,11,0.00,0.00,1.00,0.00,normal
icmp,1032,0,316,0.00,0.00,0.02,0.00,smurf
icmp,1032,0,511,0.00,0.00,0.00,0.00,smurf
```

Fig.4. Selected features of kddcup99 dataset

C. Normalization

In this step the features coming from the feature selection step are normalized. The advantage of this is accuracy increases. From literature surveyed, we analyzed that detection rate increases. We used Min-Max normalization in the proposed work.The normalization results are in fig.5.


DOS Attack Detection Based on Multivariate Correlation Analysis

```
@attribute protocol_type {tcp,udp,icmp}
@attribute src_bytes numeric
@attribute wrong_fragment numeric
@attribute count numeric
@attribute diff_srv_rate numeric
@attribute srv_diff_host_rate numeric
@attribute dst_host_same_srv_rate numeric
@attribute dst_host_serror_rate numeric
@attribute label {back,teardrop,neptune,land,smurf,pod,normal}
tcp,0.004382,0,0.013725,0,0,1,0,normal
tcp,0.004308,0,0.013725,0,0,1,0,normal
tcp,0.004015,0,0.009804,0,0,1,0,normal
tcp,0.003978,0,0.009804,0,0,1,0,normal
tcp,0.003978,0,0.009804,0,0,1,0,normal
tcp,0.003887,0,0,0,1,1,0,normal
tcp,0.002915,0,0.007843,0,0,1,0,normal
tcp,0.00385,0,0.013725,0,0,1,0,normal
tcp,0.003887,0,0.013725,0,0,1,0,normal
tcp,0.00385,0,0.033333,0,0,1,0,normal
tcp,0.003245,0,0,0,0,1,0,normal
tcp,0.00407,0,0.019608,0,0,1,0,normal
tcp,0.004594,0,0.005882,0,0,1,0,normal
```

Fig.5. Normalized features after min-max normalization

D. Triangle area generation

Triangular area is calculated in this step. Triangular area is determined using the cross product of two distinct features divided by 2.It is shown in equation (1).MCA method mentioned above is used for TAM calculation. This TAM is the correlation among features. In section IV Triangular area generation is described.

E. Mahalnobis distance calculation

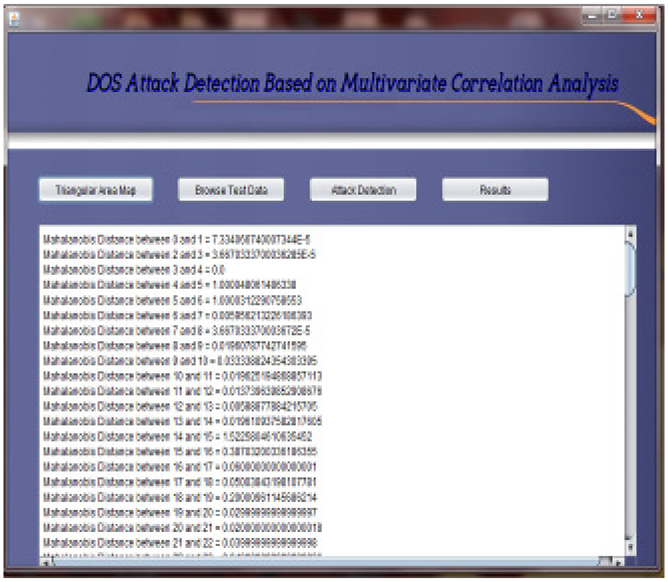
The covariance matrix is determined using the formula given below.

$$\textbf{COVARIANCE} = \begin{bmatrix} \sigma(TrA_{2,1}, TA_{2,1}) & \cdots & \sigma(TrA_{2,1}, TA_{k,k-1}) \\ \vdots & \ddots & \vdots \\ \sigma(TrA_{k,k-1}, TA_{2,1}) & \cdots & \sigma(TrA_{k,k-1}, TA_{k,k-1}) \end{bmatrix}$$

After calculating covariance mahalnobis distance is calculated. The Mahalnobis distance is calculated by using triangular area and covariance matrix .It is shown in below fig.6.

F. Classification

In the classification support vector machine classifies the attack traffic from the normal traffic.It generates SMO model by using TAM,Mahalnobis distance.SMO (sequential minimal optimization) is used in SVM classification. It separates dos type records from normal records.


DOS Attack Detection Based on Multivariate Correlation Analysis
Triangular Area Map
Browse Test Data
Attack Detection
Results

```
Mahalanobis Distance between 0 and 1 = 7.334056740007344E-5
Mahalanobis Distance between 2 and 3 = 3.667033370036285E-5
Mahalanobis Distance between 3 and 4 = 0.0
Mahalanobis Distance between 4 and 5 = 1.000048061486338
Mahalanobis Distance between 5 and 6 = 1.0000312290759553
Mahalanobis Distance between 5 and 7 = 0.005956213226108383
Mahalanobis Distance between 7 and 8 = 3.667033370003672E-5
Mahalanobis Distance between 8 and 9 = 0.019807877427415955
Mahalanobis Distance between 9 and 10 = 0.03333882435438395
Mahalanobis Distance between 10 and 11 = 0.019625164868857113
Mahalanobis Distance between 11 and 12 = 0.013738638652908679
Mahalanobis Distance between 12 and 13 = 0.005888778842157705
Mahalanobis Distance between 13 and 14 = 0.019610937582817805
Mahalanobis Distance between 14 and 15 = 1.5225984610635452
Mahalanobis Distance between 15 and 16 = 0.38703200336186355
Mahalanobis Distance between 16 and 17 = 0.04000000000000001
Mahalanobis Distance between 17 and 18 = 0.050003843198107781
Mahalanobis Distance between 18 and 19 = 0.2800006114585214
Mahalanobis Distance between 19 and 20 = 0.029999999999999997
Mahalanobis Distance between 20 and 21 = 0.020000000000000018
Mahalanobis Distance between 21 and 22 = 0.039999999999999998
```

Fig.6. Mahalanobis distance (MD)

VII. RESULTS AND ANALYSIS

In the experiment analysis we used instances of kddcup99 dataset. We used 200 instances input dataset for training purpose and 123 instances testing dataset for test purpose. The results are described in below table1.

Table1. Attack classification results

Algorithm Used	Instances Correct classified	Instances wrongly classified
Zero R	66	57
MLP	105	18
SMO(our method)	121	2

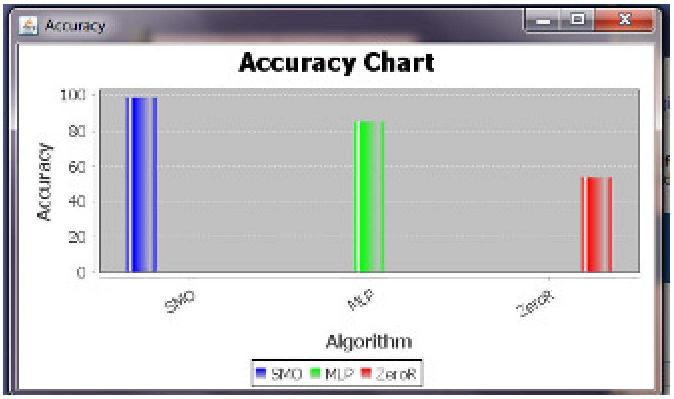


Fig.7. Accuracy comparison of smo, MLP and zeroR classifiers

The result shows that SMO (sequential minimal optimization) have better results than Multilayer Perceptron (MLP) and zero-R. MLP classifiers have good results but it takes more time than SMO. The fig.7 and fig.8 shows results. It gives more than 98 percent result for SMO. SMO requires less time as compared to MLP and ZeroR.

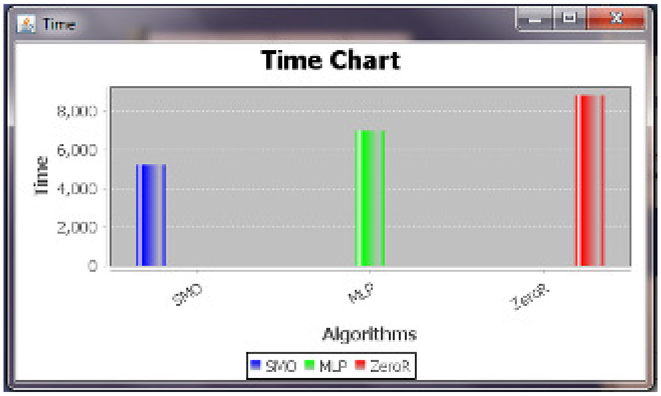


Fig.8. comparative time of smo, MLP, zeroR

VIII. CONCLUSION

The triangle area method extracts correlations between two distinct features from traffic records. Therefore proposed system uses triangle area based multivariate correlation technique (MCA). Also it uses support vector machine classification technique which classifies attacks from legitimate network traffic. The kddcup99 dataset is used for this work.The performance of this method is improved and evaluated using normalized dataset. By using the triangle area method speed up the process of multivariate correlation analysis. SMO function of support vector machine is used for the classification of attack and normal records. It gives more than 98% accuracy with less time for execution. SVM gives maximum accuracy as compared to MLP and zeroR. This method will produce better performance in terms of speed and accuracy. As future work we use real time data for this purpose.

ACKNOWLEDGMENT

I would like to express thanks to my project guide Pingale S.V. and co-guide Parihar R.B. He helped and supported for this work and No words are sufficient to express my gratitude to my family for their unwavering encouragement. I also thanks to all my friends for being a constant source of my support.

REFERENCES

[1] Shuyuan Jin and Daniel S. Yeung ,"A Covariance Analysis Model for DDoS Attack Detection" IEEE **2004.**

[2] Gloria C.Y. Tsang, Patrick P.K. Chan, Daneil S. Yeung and Eric C.C. Tsang "Denial of service detection by support vector machines and radial-basis function neural network" IEEE ***2*****004.**

[3] Venkata Suneetha Takkellapati and G.V.S.N.R.V Prasad," Network Intrusion Detection system based on Feature Selection and Triangle area Support Vector Machine" International Journal of Engineering Trends and Technology- Volume**3,** Issue**4**- **2012**

[4] Chih-Fong Tsai and Chia-Ying Lin, "A triangle area based nearest neighbors approach to intrusion detection" Pattern Recognition, vol.**43**, pp. **222 – 229, (2010)**.

[5] Wei Wang , Xiangliang Zhang , Sylvain Gombault , and Svein J. Knapskog , "Attribute Normalization in Network Intrusion Detection" 10th International Symposium on Pervasive Systems, Algorithms, and Networks, **2009**.

[6] Zhiyuan Tan1, Aruna Jamdagni, Xiangjian He, Priyadarsi Nanda, and Ren Ping Liu," Multivariate Correlation Analysis Technique Based on Euclidean Distance Map for Network Traffic Characterization" Research Centre for Innovation in IT Services and Applications (iNEXT)

[7] B. Kiranmai and A. Damodaram, "A Comprehensive Survey on Methods Implemented For Intruder Detection System" International Journal of Computer Sciences and Engineering volume-**2**,Issue-**8**,E-ISSN:**2347-2693**

[8] Zhiyuan Tan,Aruna Jamdagni,Xiangjian He,Priyadarshi Nanda and Ren Ping Liu, "A System for Denial of Service Attack detection based on Multivariate Correlation Analysis" IEEE Transaction on parallel and distributed systems,vol.**25**,No.**2**,February **2014**

[9] Peter Harrington,"Machine Learning in Action", Manning Publications ©**2012**.

[10] Wun-Hwa Chen, Sheng-Hsun Hsu, Hwang-Pin Shen "Application of SVM and ANN for intrusion detection" Computers & Operations Research 32 (**2005**) **2617–2634**