# A system based on Naive Bayesian for Denial-Of-Service Attack detection

Sana Chaugale
Computer Engineering
G. H. Raisoni College of Engineering and Management
Wagholi, Pune, India.
*Sanna.chaugale@gmail.com*

Dr. Tanuja Dhope
Computer Engineering
G. H. Raisoni College of Engineering and Management
Wagholi, Pune, India.
*tanuja.dhope@raisoni.net*

***Abstract*—**Denial-of-service (DoS) attacks cause serious effect on systems. For most correct network traffic characterization, attack detection system uses multivariate correlation analysis (MCA). It Extract the geometrical correlations in between network traffic features. MCA based system enlightens the principle of anomaly based detection while attack recognition. MCA makes the situation easy for detecting known and unknown types of DoS attacks by simply observing the legitimate network traffic patterns. MCA uses Triangle Area Map (TAM) technique to speed up the Multivariate Correlation Analysis process. Proposed system can be evaluated by using KDD cup99 dataset. Naive Bayes (NBS) classifier is used as for attack detection. This algorithm addresses the problem of classifying the large intrusion detection dataset, which improves the detection rates and reduces the false positives at acceptable level in intrusion detection.It is probabilistic classifier which based on applying Bayes theorem.The proposed DoS attack detection system achieved highest accuracy as comparing to RBFN and IBK.99.96% accuracy is achieved by intrusion detection system.The Proposed detection system gives very low false positive Rate as about 0.002% which helps to increase the performance of detection System. As compare to RBFN and IBK, Naïve bayes classifier gives very low false positive rate, which helps to increase the performance of detection System. As compare to RBFN and IBK, Naïve bayes classifier gives very low false positive rate.

***Keywords****-Denial of service, KDD cup99 dataset, Multivariate correlation analysis, triangle area, Naive bayes classification*

*****

## I. INTRODUCTION

Denial of service attack degrades the efficiency of online services. Therefore it is more important to detect DoS attack to prevent online services. The DoS attack detection system, focuses mainly on the network based detection mechanism. The detection system includes two approaches, misuse detection and anomaly detection. For the identification of known attack Misuse detection is used, by using the signatures of predefined rules. Anomaly detection is used to establish the usage profile of the system. After the detection there are two phases named training and testing phase. During the training phase, the profiles for the legitimate traffic records are generated and the generated records are stored in the database. The trusted profile generation is build and then it is given to “attack detection” module, which compares the individual tested profile with the normal profile.

Multivariate correlation analysis discovers the relations among features within the observed data objects. Significant changes of these relations indicate occurrences of intrusions within observed data objects. It achieves high detection accuracy while retaining a low false positive rate. Moreover, benefiting from the principal of anomaly detection, DoS attack detection approach is independent on prior knowledge of attack and is capable of detecting both known and unknown DoS attacks. Triangle area based technique is used to increase speed of the multivariate correlation analysis. System is valued using KDD Cup knowledge.

Naive Bayesian classifiers assume that the effect of an attribute value on a given class is independent of the values of the other attributes. This assumption is called class conditional independence. It is made to simplify the computations involved and, in this sense, is consider “Naive”. Naive Bayesian classifiers allow the representation of dependencies among subsets of attributes [1]. Though the use of Bayesian networks has proved to be effective in certain situations, the results obtained are highly dependent on the assumptions about the behaviour of the target system, and so a deviation in these hypotheses leads to detection errors, attributable to the model considered [2].A Naive Bayes classifier is used in attack detection module, Assumption of Naive Bayes classifier is that the presence or absence of a particular feature in a class is totally unrelated to the presence or absence of other feature within that class. A DoS detection system is evaluated by using KDD Cup 99 dataset and plays the role of state of the art system .It improves detection rate performance for various types of intrusions within network. The main purpose here is to improve the performance of Naive Bayesian classifier for intrusion detection and minimize false positive rate.

## II. LITERATURE SURVEY

In paper [3], various types of denial of service attacks and the effect of Dos attack on the performance discussed. In this paper several intrusion detection techniques are studied. All jamming techniques and the algorithms for Dos detection, throughput of this algorithm reduces the performance of the network. In [4] paper, shown, KDD’99 dataset that is used in Dos attack detection. Most of anomaly detection methods evaluation is done by using KDD’99 dataset [5] . DARPA’98 is compressed raw (binary) tcp dump data of 7 weeks of network traffic of about 4 gigabytes and is built on the bases of data capturing in DARPA’98 IDS evaluation [6].In [7] paper, Discussed Multivariate Correlation Analysis(MCA). MCA extract the correlation between the features observed in traffic record. Triangle area Map generation (TAM) is purposely used to make fast multivariate correlation analysis. In paper [8], introduced the technique of game theory, this technique provides powerful tools to model and analyse such attacks it discussed about MAC layers jamming games among the set of jammers and transmitters. Result comes out from this game theory provide robust network protocol design for secure wireless communication and characterize the expected performance under DoS attacks. In [9] multiple applications, parameter estimation for Naive Bayes models use the method of similarity, It means any one can work with Bayesian model without considering the probability and any Bayesian method. To estimate the parameters like means and variances of the variables it requires a very few training data and it is main advantage of the naive Bayes classifier. It is not necessary to determine whole covariance matrix only variable variances are under consideration, because only independent variables are assumed here.

## III. PROPOSED SYSTEM

A dos attack detection system uses multivariate correlation analysis that is used for accurate network traffic characterization. Correlation analysis extracts the geometrical correlations between legitimate network traffic features. Dos attack detection system by using multivariate correlation analysis considers the principle of anomaly detection in attack recognition. Only with the learning of network traffic patterns it is very easy to detect different known and unknown dos attack. To improve the effectiveness of proposed detection system KDD cup 99 dataset is used. Furthermore, a triangle-area-based technique is used to enhance and to fast up the process of multivariate correlation analysis. All the extracted correlation are stored in a place called “triangle area map”(tam), are then used to replace the original records or normalized feature record to represent the traffic record. It does differentiate in between legitimate and illegitimate network traffic records.

The anomaly detection mechanism is adopted in decision making. A bayes classifier is a simple probabilistic classifier based on bayes theorem. In simple way, a naive bayes classifier assumes presence or absence of a particular feature of a class. For example, a fruit may be an apple if it is red in color, round in figure, and about four inches by diameter. These features are depends upon each other a naive bayes classifier consider all of these features are independently to contribute the probability that the give fruit is
an apple. Depends on the detail nature of the probability model, Bayesian classifier can be trained efficiency

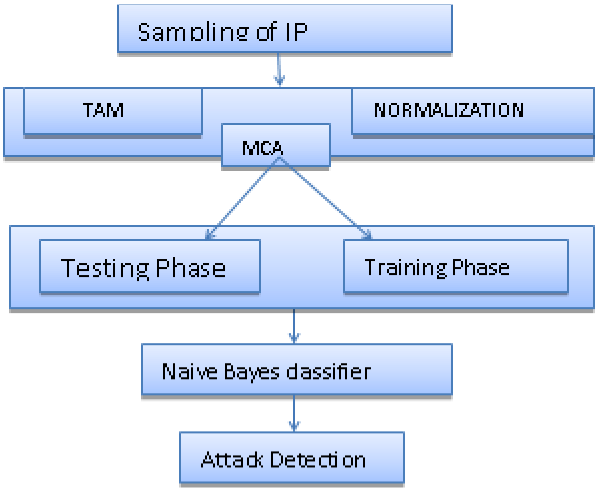


Figure1. DoS Attack Detection Using Naive Bayes Classifier

## IV. ATTACK DETECTION

### A. Algorithm For Normal Profile Generation Based on Triangle Area

```
1.$X_{TAM_{lower}}^{normal}$ with set of n legitimate training traffic record
2. $\overline{TAM_{lower}^{normal}} \leftarrow \frac{1}{n}\sum_{i=1}^{n} TAM_{lower}^{normal\ ,i}$
3. Generate Covariance Matrix ($Cov$) for each traffic record
4. For each i=1 to n go through next
5. Compute distance between $TAM_{lower}^{normal\ ,i}$ $and\overline{TAM_{lower}^{normal}}$
6. End for
7. $\mu \leftarrow \frac{1}{n}\sum_{i=1}^{n} MD^{normal\ ,i}$
8.$\sigma \leftarrow \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(MD^{normal\ ,i} - \mu)^2}$
9. Normal Profile (Pro) ← (N ($\mu,\sigma^2$),$\overline{TAM_{lower}^{normal}}$ ,$Cov$)
10. Return Pro (Normal profile)
```

### B. Algorithm For DoS Attack Detection

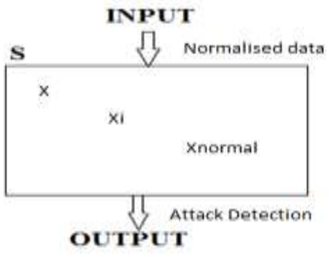
```
1: for each observed traffic record $X^{obsereved}$ Generate $TAM_{lower}^{obsereved}$
2:$MD^{obsereved}$ ← MD ($TAM_{lower}^{observed}$ , $\overline{TAM_{lower}^{normal}}$
3: if $(\mu - \sigma *\alpha) \leq MD^{obsereved} \leq (\mu + \sigma *\alpha)$ then
4: Do classification by using Naive Bayes Classifier
5: return Normal
6: else return an attack
7: end
```

### C. Naive Bayes Classifier Algorithm

```
Step 1: Input dataset called (Intrusion Detection Dataset.)
Step 2: Pre-process dataset.
Step 3: Alert graph model generation.
Step 4: Alert generation using Naive Bayes classification.
Step 5: Compute Detection Rate (DR) and False Positive Rate
Step 6: Output as detected intrusion.
```

### D. Mathematical Model

**Input:** Normalized Data X={x1, x2, x3....xn}



**Output:** Attack Detection for Individual Record
Xi= {F1, F2, F3 ...Fn}

### E. Equations

probabilities of correctly classifying a sample into its distribution using the sample-by-sample labeling as the cumulative distribution functions shown in (1) and (2), respectively.

$$p1 = \int_{-\infty}^{\overline{\mu}} \frac{1}{\sigma 1\sqrt{2\Pi}} e^{-(x-\mu_1)^2\ /2\sigma^2{}_1} \mathrm{dx} (1)$$

$$\mathrm{P2}= \int_{\overline{\mu}}^{+\infty} \frac{1}{\sigma 2\sqrt{2\Pi}} e^{-(x-\mu_2)^2\ /2\sigma^2{}_2} \ \mathrm{dx} \quad (2)$$

Where,

$$\bar{\mu} = \mu_1 \times \frac{\sigma_2}{\sigma_{1+\sigma_2}} + \mu_2 \times \frac{\sigma_1}{\sigma_{1+\sigma_2}}$$

$\bar{\mu}$ is threshold value for classifying the sample P1 and P2 defines probability distributions.

The probability of correctly classifying all k samples is

P r (k) = $P^k$(3)
k = mean of random samples

*1)Normal Profile Generation:*

Xnormal={X1normal, X2normal, Xn normal} is normalized data.

The generated lower triangles of the set of g traffic records are denoted by

$X_{TAMlower}^{normal} = \{ TAM_{TAMlower}^{normal\ ,1} , TAM_{TAMlower}^{normal\ ,2} ,\ldots\ldots\ldots, TAM_{TAMlower}^{normal\ ,g} \}$
(4)
Threshold is calculated by ,The threshold given is used to differentiate attack traffic from the legitimate one,

Threshold = μ + $\sigma$ * $\alpha$(5)
The covariance between two arbitrary elements in the lower triangle of a normal TAM is

$$Cov=\begin{bmatrix} \sigma(Tr_{2,1}^{normal}, Tr_{2,1}^{normal}) & \cdots & \sigma(Tr_{2,1}^{normal}, Tr_{m,m-1}^{normal}) \\ \vdots & \ddots & \vdots \\ \sigma(Tr_{m,m-1}^{normal}, Tr_{2,1}^{normal}) & \cdots & \sigma(Tr_{m,m-1}^{normal}, Tr_{m,m-1}^{normal}) \end{bmatrix}$$

The covariance between two arbitrary elements in the lower triangle of a normal TAM is

$\sigma(Tr_{j,k}^{normal}, Tr_{l,v}^{normal}) = \frac{1}{n-1}\sum_{i=1}^{n}\left(Tr_{j,k}^{normal\ ,i} - \mu Tr_{j,k}^{normal}\right) \times (Tr_{l,v}^{normal\ ,i} - \mu Trl,vnormal)$ (6)
The threshold given is used to differentiate attack traffic from the legitimate oneNaive Bayes classification works according to assumption is as,

$$P\ (c_j|d) = \frac{p(d|c_j)\ p(c_j)}{p(d)} \qquad (7)$$

Where,
P ($c_j$| d) = probability of instance *d* being in class *cj*, this is that we have to compute
P (d | $c_j$) = probability to generate d instance of class $c_j$, consider that being in class $c_j$, affects to have class feature d with some given probability
P ( $c_j$ ) = probability that class cj occurs; this shows the frequency of the class $c_j$.
P (d) = probability that distanceoccur this can ignored, since it is mostly same for all the classes.
Naïve bayes classifier considers that attributes are independent to make task too simple.
P (d$c_j$) = p (d1|$c_j$) * p (d2|cj) * ….* p (dn$c_j$)
Where,
P (d|$c_j$) = Probability that class $c_j$ generate d instance.
P (d1|$c_j$ = Probability that class $c_j$ generate the observed value for the first feature.
P (d2|$c_j$ = Probability that class$c_j$ generate the observed value of second feature of class.
P (dn$c_j$) = Probability that class cj generate the observed value for nth feature.

## V. Results

The detection performance of naivy bayes classifier according to time is given below. Naive bayes takes low time to classify a DoS attack as compare to RBFN and IBK classifier.

Time measured in milliseconds. The results of RBFN and IBK are calculate by providing data set to weka while results according to naive bayes are generated by implementing it, then after comparison of Naïve Bayes,RBFN,IBK held out.

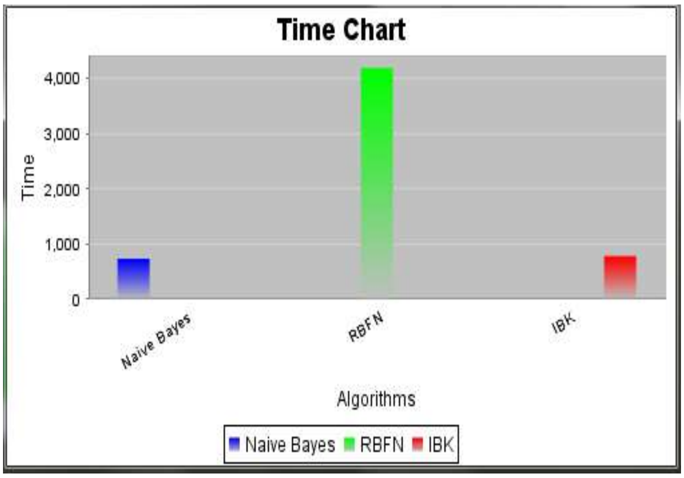


Figure2. Time graph of Detection System

The performance of naivy bayes classifier according to Accuracy is given below

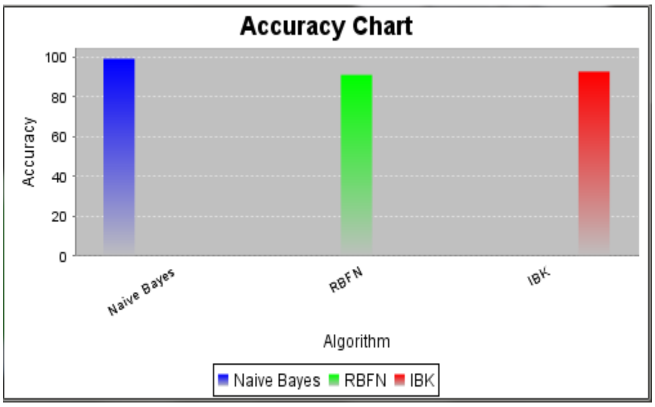


Figure3.Accuracy performance

The proposed system achieved highest accuracy as comparing to RBFN and IBK.99.96% accuracy is achieved by intrusion detection system.

The Proposed detection system gives very low False positive Rate as about 0.002% which helps to increase the performance of detection System. As compare to RBFN and IBK, Naïve bayes classifier gives very low false positive rate.

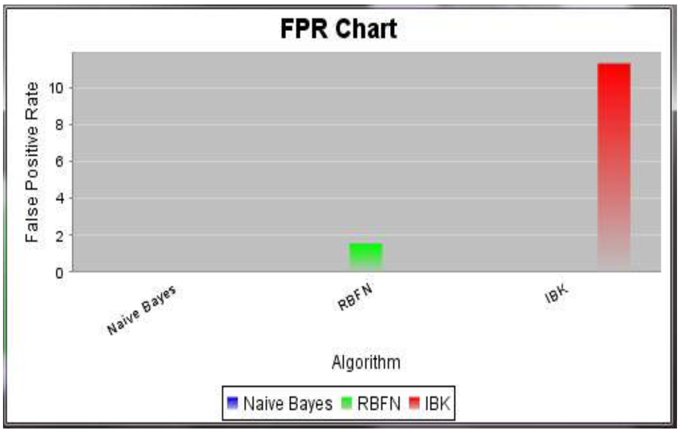


Figure4. False positive Rate (FPR) given by proposed system

As compare to RBFN and IBK, Naïve bayes classifier gives very low false positive rate.

## VI. CONCLUSION

This paper presented a “Multivariate Correlation Analysis” for DoS attack detection system which uses the triangle area technique for anomaly-based attack detection technique. It extracts the correlation in between two distinct features within each traffic record. Extracted correlations are then stored in TAM map. MCA use feature normalization technique. This provides very important information that is useful for differentiating between legitimate and attacked one traffic.

To verify the effectiveness and to improve the performance the denial of service attack detection system KDD cup 99 dataset and Bayes classifier is used. KDD cup 99 dataset is publicly available dataset. It has been widely used for intrusion detection approaches.

Intrusion detection using naive Bayesian classifier is suitable for analyzing large number of network logs or audit data. Bayesian classifier work with probability and it improves the performance of detection rates for different types of intrusions. The main propose of naive Bayes classifier is to improve the performance and fast classification of attack.

## REFERENCES

[1] Manish Jain, Prof. Vineet Richariya, “An Improved Techniques Based on Naive Bayesian for Attack Detection,” International Journal of Emerging Technology and Advanced Engineering,2012.

[2] Proano, A. Lazos, “Packet-Hiding Methods for Preventing Selective Jamming Attacks”, Dependable and Secure Computing, IEEE, 2011.

[3] Konstantinos Pelechrinis, Marios Iliofotou and Srikanth V. Krishnamurthy, “Denial of Service Attacks in Wireless Networksthe Case of Jammers”, Communications Surveys & Tutorials, IEEE, 2011.

[4] “Nsl-kdd data set for network-based intrusion detection systems.” Available on: http://nsl.cs.unb.ca/NSL-KDD/, March 2009.

[5] Sagduyu, Y.E.Berry, R.A.Ephremides, A.” Jamming games in wireless networks with incomplete information”, Communications Magazine, IEEE, 2011

[6] J.Welkin Eyes, S.Karthiprem, E.Thangadurai “High Accuracy detection of Denial of Service Attack based on Triangle Map Generation”IJCSMC, 2014.

[7] Ghosal, A. Halder, S. Mobashir, M. Saraogi, R.K. Das Bit, S.” A jamming defending data forwarding schemefor delay sensitive applications in WSN” Wireless Communication, Vehicular Technology, Information Theory and Aerospace &ElectronicsSystems Technology,”2nd International Conference, 2011.

[8] K.Sujithra1, V.Vinoth Kumar “A Survey on Triangle Area Map Based Multivariate Correlation Analysis to Detect Denial-Of Service Attack” International Journal of advanced research in computer and Communication Engineering Vol. 3, Issue 10, October 2014

[9] Zhiyuan Tan, Aruna Jamdagni, Xiangjian, Priyadarsi Nanda, and Ren Ping Liu,”A System for Denial-of-Service Attack Detection Based Multivariate Correlation Analysis,” IEEE transactions on parallel and distributed systems 2014

[10] Z. Muda, W. Yassin, M.N. Sulaiman, N.I. Udzir. “Intrusion Detection based on K-Means Clusteringand Naive Bayes Classification,” 7th International Conference on IT in Asia, 2011.

[11] 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,” IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications, Liverpool, UnitedKingdom,2012.