# Attack Detection within Network using Hybridization of Bisection TDE

**Amit Dogra , Taqdir Abstract**

*Abstract: Today the leading cause of attack is traffic distribution that affects transmission as well as network performance. Traditionally the mechanism is utilized for detecting traffic distribution is not so accurate, user friendly and also they are time consuming. Therefore in this paper a hybrid approach is used that detect attacks by eliminating replicated features. It utilizes TDE with bisection mechanism that has great attack detection rate  The result of the proposed system as compared to existing system is better by the margin of 8 to 10%.. The primary reason for the betterment is because of contrast enhancement and scaling factor at pre-processing and objective function overhead consideration at segmentation and classification phase.  The results are shown in terms of accuracy, PSNR and MSE.*

*Keyword: TDE method, PSNR, traffic distribution.*

## I. INTRODUCTION

The main control centre of network is bandwidth that manage the overall activities of network like movement , memory , traffic etc. there are various disabilities that affect bandwidth activities . These are known as traffic or bulk attack. The attacks damage the bandwidth and also affect the network's memory and other daily life's activity.  The various diagnostic methods are used for detecting traffic distribution and it utilizes mechanism to improve attack detection. But the current methods are not user friendly and invasive so there is need to develop convenient , non invasive and user friendly detection and diagnostic system of traffic distribution. The Traditional differential technique is utilized for this purpose. Data segmentation is the process in which data is partitioned into multiple regions which results into number of fragments of objects. The main purpose of data segmentation is to make modifications and simplifications in data to make it more clear and meaningful. The main feature of segmentation is to find boundaries and objects in data .Resultant of data segmentation composed of multiple regions that entirely cover the whole data. Data in a common area are alike to each other on the basis of some computed property such as common difference, frequancy. It is computer based process of bringing out useful information from the data. Input is the data and process ends with an output that is data which is numerical. Data processing is a technique by which we can change the input data by applying   mathematical algorithms to produce an output data that is enhanced.

  **Dr Taqdir\*,** Assistant Professor   in the department of Computer Science and Engineering at Guru Nanak Dev University ,R/C Gurdaspur.
  **Amit Dogra,** Assistant   Professor in the department of computer Science and Engineering in SoET,BGSB University Rajouri (J&K).

## II. LITERATURE SURVEY

**Ms. Pooja S. Deshpande1, Prof. S.J. Honade**, 2017 presented [1] a firefly algorithm for detection and segmentation of traffic in the bandwidth using UCI data, this is the algorithm which work in an sequence manner from traffic to attack. Firstly, the area of the whole traffic is segmented and then the abnormal attacks are separated based on obtained area of traffic. The Firefly Algorithm has provide us with a better parameters   like delay time (0.0195214 Sec.), traffic  percentage is (1.4016429%),  this is used for diagnosis of traffic  using UCI data. Therefore the firefly algorithm has provide  with  a better and clear results as compared to linear SVM algorithm , Radial basis SVM and Quadratic SVM algorithm as the region that is segmented is better and delay time is also less. So, the FA is the better .

**Emrah Hancer1, Celal Ozturk2 and Dervis Karaboga,2013 presented[2]**    Artificial bee colony algorithm for detection and segmentation of traffic in the bandwidth using UCI data. This proposed methodology works on 3 principles .Firstly, noise is removed from and data is enhanced . Secondly data is processed using ABC and in last final post processing is performed to extract attacks. The proposed methodology has been compared with K-means, GA, FCM based data segmentation methodologies.  The ABC has a powerful numerical and visual outputs for diagnosis of traffic.

**Abdalla Mostafa et.al, 2015 presented[3]** Artificial bee colony algorithm for attack segmentation. Correct attack segmentation plays a vital role in computer aided diagnosis (CAD) system. Cad is composed of 3 phases that is pre processing ,segmentation and classification. This paper has proposed a attack segmentation by using clustering process that is boost up by ABC. In preprocessed phase   data annotations are removed and ribs are joined, ABC is applied for having initial segmentation of attack. At last, refining is done by region growing approach to have final segmented output. Experimental results obtained from 38 ATTACK datas demonstrate excellent segmentation.

**Tinali Kamble, PrachiRane, 2013 presented[4],** proposed swarm intelligence approach for  attacks detection . Previous work done in the field of traffic detection and segmentation have also discussed in this paper. This   methodology provide us with the better visual results of   attacks segmentation .the proposed work further can be continued with classification of  attacks by applying artificial neural network.

**M.Mary Synthuja Jain Preetha et.al, presented[5]**, an algorithm for common difference data segmentation that is totally based on region growing. Basicaly it has phases, In first phase selection on initial seeds are done by having analysis of histogram than regions are allowed to grown from these selected seeds. In next phase threshold is optimized again but the use of cuckoo search algorithm. At last results of both has been compared against various evaluation parameters.

**Xiaolin Zhang, Tao Yang et.al ,2015 presented[6],** the limitations of older technique that is 2d entropy threshold segmentation. The researcher has proposed a flame data segmentation method that utilize artificial bee colony algorithm having the properties of levy flights. By utilizing strategy of levy flight id creates a recently created space of local search and increase the convergence rate. Computational results provide us with a methodology that is powerful in means of threshold selection correctness and reduced run time.

### III. PROPOSED METHODOLOGY

In the proposed system we are using two classifiers namely bounding box and Haar like mechanism. The bounding box is used for detecting the region of interest and then Haar Like mechanism is used to detect feature from region of interest.

#### A. Bisection TDE Mechanism

In this mechanism the box surrounds the data in the training data with K dimensional and then compute the vector V using various frequency in that box.

$$y = \beta + \sum_i \alpha_i(V, V_i) \qquad (1)$$

Here y is coordinate , $\beta, \alpha$ are constant values, V is vector for various frequency

#### B. Haar Mechanism

In this the data is firstly decomposed using 2D Haar mechanism in multiple levels. The process produces four coefficients, i.e. approximation, horizontal, vertical and diagonal coefficients. Among these four coefficients, only approximation coefficient is chosen to generate feature vectors. The equations are as given below:

$$L_K(M) = \iint K(M - x)I(x)dx \qquad (2)$$

Here K represents coefficient , M denotes the chosen feature, I(x) represents the number of feature vector.

*Flowchart*
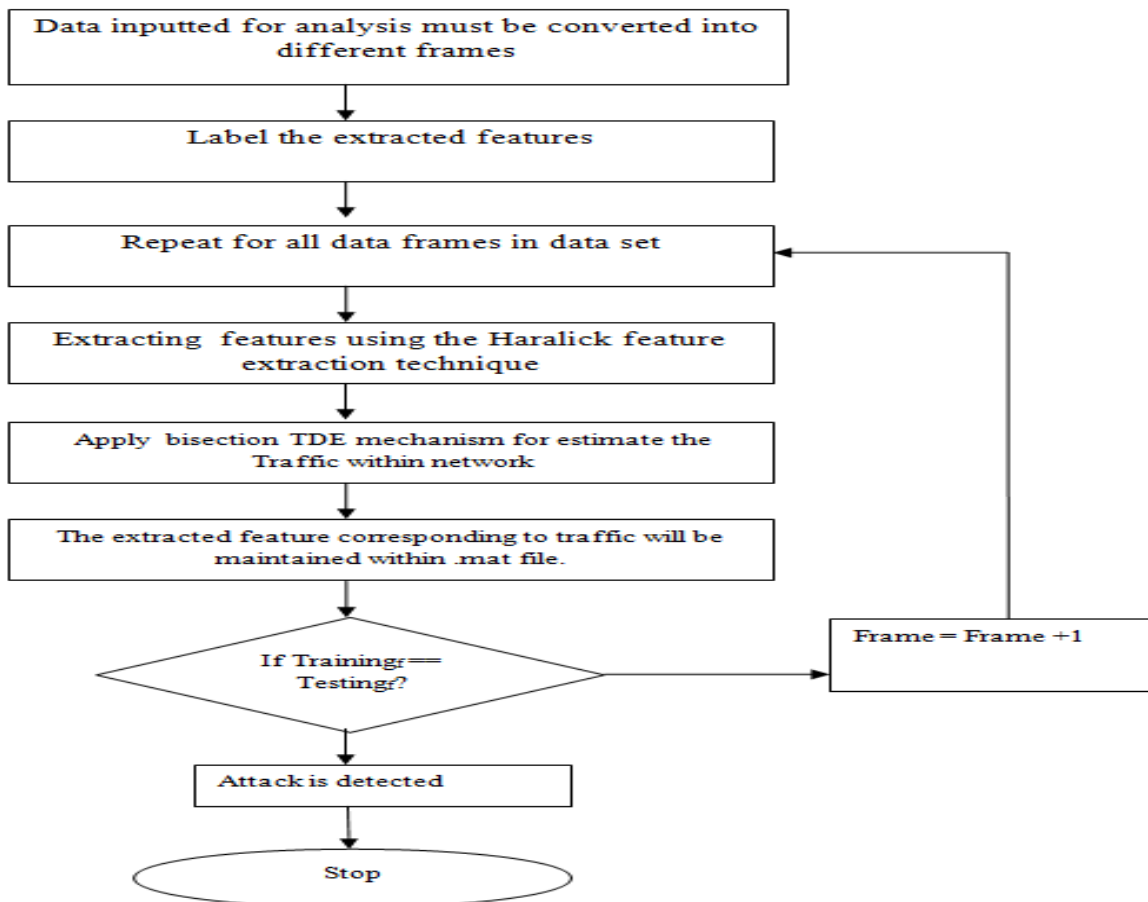
### A.  Algorithm

**Training**

- Perform training using Haralick and Bisection mechanism from training datas. (These features define region of interest (ROI) that are inputted by the user looking for an object)

**Testing**

- First of all, datas are inputted for analysis must be converted into data frames. The standardized library of MATLAB for signal analysis can be used for this purpose. Mmread function is available in MATLAB and format of processing is .csvi.
- After extracting the data, each data is labelled as Data1, Data2, Data3, …, Datan
- Loop from first data to last data is applied for processing each data frame
  a) Extracting features using the Haralick feature extraction technique and store it in .mat file (Assuming frequency features of traffic data is different)
  b) Multiheuristic bisection mechanism can be used to estimate the height of the object from the data
  c) The extracted feature corresponding to height will be maintained within .mat file.
  d) If $(Training_f == Testing_f)\ then$
     Object detected
     Return
     Else
     Frame=Frame+1
     End of if

  End of Loop

- Stop

This methodology correctly classifies leaf into distinct categories. The region of interest is increased through the H-Bisection TDE classifier and hence more region is classifier and non-segmented region is reduced. Moreover, proposed approach is dynamic since any dataset can be analyzed rather than fixed set of datas hence applications of both supervised and unsupervised learning.

## IV.  RESULTS

The result obtain after applying bisective TDE mechanism is in terms of PSNR and MSE. In addition classification accuracy is also considered for proving validity of study.

This metric must be enhanced and indicates correctly identified segments from the datas within the dataset. Classification accuracy is expressed in the form of true positive, true negative, false positive and false negative parameters as in the equation 3

$$Classification_{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Equation 8

TP is the true positive and TN is the true negative values obtained during simulation.

**Table.1: Classification Accuracy**

| Used Data | Classified as | TDE Accuracy (in %) | TDE-B Classifier (in %) |
|---|---|---|---|
| 1000 tuples | Attack Detected | 92.3 | 96 |
| 2000 tuples | Attack Not Detected | 94 | 97 |
| 3000 tuples | Attack Detected | 90 | 95 |

Classification accuracy is identified in terms of correctly classified segments from the data dataset. Result is expressed in the form of true positive rate. True negative rate is not observed while considering this parameter and hence classes identified may vary with time. In other words, classification accuracy for all the categories may not be symmetrical.

The peak signal to noise ratio obtained from the proposed system is better by 2 to 3%. The existing system PSNR is in the range of 12-13 where as proposed system peak signal to noise ratio is between 16-21. The peak signal to noise ratio must be higher for higher convergence rate. the evaluation equation for peak signal to noise ratio is given as

$$PSNR=10\log_{10}\left(\frac{MSE^2}{255}\right)$$

Equation 9: Peak signal to noise ratio

Peak signal to noise ratio is expressed as mean square error and 255 is the maximum intensity of pixel. Table 2 gives the peak signal to noise ratio of existing and proposed system.

**Table 2: PSNR of existing and proposed system**

| Used Data | Classified as | TDE PSNR | TDE-B Classifier PSNR |
|---|---|---|---|
| 1000 tuples | Cancerous | 12.3 | 16 |
| 2000 tuples | Cancer Not Detected | 12.4 | 17 |
| 3000 tuples | Cancer Detected | 12.6 | 18 |

The peak signal to noise ratio expressing the desired levels of accuracy from the proposed system is better as compared to existing system. The plot elaborating table 6.3.2 is given as The mean square error is another parameter on which classification accuracy is greatly dependent.

Peak signal to noise ratio goes side by side along with mean square error. Higher the PSNR least will be MSE. In other words, mean square error can be obtained from PSNR itself.

$$MSE = \sum_{i=1}^{n}(1 - 10 * \frac{\log(X_i)}{255})/n$$

Equation 10: Mean square error evaluation

'X' is the intensity of the pixel present within the data. 'n' is the total number of data within the data. As the mean square error decreases in the proposed system since population decreases, hence execution time also decreases subsequently.

**Table .3: Execution time of TDE and bisection TDE**

| Used Data | Classified as | TDE CPU time | TDE-B Classifier CPU time |
|-----------|---------------|--------------|---------------------------|
| 1000 tuples | Attack detected | 1.3098 | 1.007 |
| 2000 tuples | Not Detected | 1.4562 | 1.095 |
| 3000 tuples | Attack Detected | 1.8786 | 1.0083 |

Execution time reduction is due to feature vector reduction and replicated features are eliminated from the simulation. The population size decreases and hence overall execution time is decreases significantly.

## V. CONCLUSION

The hybrid approach of TDE with bisection mechanism used in the proposed system eliminate the replicated features and thus malicious data is detected at great rate. The result of the proposed system as compared to existing system is better by the margin of 8 to 10%.. The primary reason for the betterment is because of contrast enhancement and scaling factor at pre-processing and objective function overhead consideration at segmentation and classification phase. Since we are working upon the networking field, so quick result generation is the need of the hour. This is achieved using the proposed system. Individual distinct features obtained can be used for real time data also for attack detection. In future this mechanism can be tested against the real time data and performance can be tested on GPU with deep learning.

## REFERENCES

1. http://www.iosrjournals.org/iosrjece/papers/Vol.%2012%20Issue%202/Version-3/T120203129144.pdf
2. Hancer, E., Ozturk, C., & Karaboga, D. (2013, November). Extraction of attackss from UCI datas with artificial bee colony based segmentation methodology. In Electrical and Electronics Engineering (ELECO), 2013 8th International Conference on (pp. 516-520). IEEE.
3. Mostafa, A., Fouad, A., Elfattah, M. A., Hassanien, A. E., Hefny, H., Zhu, S. Y., & Schaefer, G. (2015). CT attack segmentation using artificial bee colony optimisation. Procedia Computer Science, 60, 1622-1630.
4. Kamble, T., & Rane, P. (2013). Attacks segmentation using swarm intelligence approach. International Journal of Scientific & Engineering Research, 4(5).
5. Preetha, M. M. S. J., Suresh, L. P., & Bosco, M. J. (2016). Region based data segmentation using cuckoo search algorithm. Journal of Chemical and Pharmaceutical Sciences, 9(2), 884-888.
6. Zhang, X., Yang, T., & Cui, N. (2015). Flame data segmentation based on the bee colony algorithm with characteristics of Levy flights. Mathematical Problems in Engineering, 2015.
7. Pathak, S., & Sejwar, V. (2017). A Review on Data Segmentation Using Different Optimization Techniques. International Journal of Computer Sciences and Engineering, 5(5).
8. Cuevas, E., Sención-Echauri, F., Zaldivar, D., & Pérez, M. (2013). Data segmentation using artificial Bee colony optimization. In Handbook of Optimization (pp. 965-990). Springer, Berlin, Heidelberg.
9. Singh, V., & Misra, A. K. (2017). Detection of plant leaf attacks using data segmentation and soft computing techniques. Information Processing in Agriculture, 4(1), 41-49.
10. Singh,I,& Gupta, N. (2015). Optimized Attack Segmentation using Ant Colony Optimization. Methodology, 4(9).
11. https://zapdoc.tips/a-survey-on-multi-level-thresholding-for-data-segmentation.html
12. Sivaramakrishnan, A., & Karnan, M. (2014). Network data segmentation using firefly algorithm and enhanced bee colony optimization. In International conference on information and data processing (ICIIP-2014) (pp. 316-321).

## AUTHORS PROFILE

**Dr.Taqdir,**
Designation : Assistant Professor
Qualification : PhD(CSE,NIT Jalandhar),M.Tech(CSE,),B.Tech(CSE)
Department : Computer Science And Engineering
Email : taqdir_8@rediffmail.com
Dr Taqdir is currently working as Assistant Professor in the department of Computer Science and Engineering at Guru Nanak Dev University ,R/C Gurdaspur. She hold the experience of 17 years in teaching at University level and her area of interest are Digital Image Processing, Machine Learning.

**Mr. Amit Dogra,**
Designation : Assistant Professor
Qualification : M.Tech, B.E
Department : Computer Science And Engineering
Email : amitdogra004@gmail.com.
Amit Dogra is currently working as Assistant Professor in the department of computer Science and Engineering in SoET,BGSB University Rajouri (J&K). He hold B.E degree in Information Technology and M.Tech degree in Networking and Internet Engineering.