DDOS ATTACK DETECTION ON CLOUD ENVIRONMENT GENATIC ALGORITHM AND FUSSY ESTIMATOR METHOD ON WIRELESS SENSOR NETWORK WITH DARPA2000 AND CADIA2008 DATASET

Amarjeet kaur Er. Jashanpreet kaur
M.Tech Research scholar Assistant Professor
Guru Kashi University , Talwandi Sabo

Abstract: The DoS attack strategy is depend on sending many harmful packets to the victim system or device this cause overloaded and resource consuming . The DDoS strategy is depends on start generating as many packets as they can toward the victim. Systems used to detect DDoS attacks are considered as a type of Network Intrusion Detection Systems (NIDs). The latter employ two distinct approaches to detect malicious activities: signature-based detection and anomaly-based detection. In the recent decade, many anomaly-based detection methods were proposed to identify DDoS attacks from network traffic. Basically, these detection methods can be classified into two categories: off-line DDoS mining and on-line DDoS detection. Off-line DDoS mining usually try to find attacks by analyzing the main characteristics of feature distributions of the network traffic with some systematic methods, such as PCA (Principal Component Analysis)) and dominate states analysis. An important such perspective in terms of detecting DoS attacks is to view the problem as that of a classification problem on network state (and not on individual packets or other units) by modelling normal and attack traffic and classifying the current state of the network as good or bad, thereby detecting attacks when they happen. To propose DDOS Attack Detection On Cloud Environment Genetic algorithm and fuzzy estimator method on wireless sensor network with DARPA2000, CAIDA2007 and CADIA2008 dataset. In this work 94% accuracy is achieved on DARPA dataset with the help of Fuzzy logic classifier.

Keywords: DDOS, Attacks, NID, cloud, network , dataset etc.

I.INTRODUCTION

Nowadays, distributed denial of service (DDoS) attacks pose one of the most serious security threats to the Internet [1]. DDoS attacks can result in a great damage to the network service. The DDoS attackers usually utilize a large number of puppet machines to launch attacks against one or more targets, which can exhaust the resources of the victim side. That makes the victim lose the capability to serve legitimate customers and prevent legitimate users from accessing information or services. Since DDoS attacks can greatly degrade the performance of the network and are difficult to detect, they have become one of the most serious security challenges to the current intrusion detection systems (IDS) [2]. Concerning the current state of the network, every corner of the world is likely to be the target of DDoS attacks. However, as long as they are detected early, the loss can be reduced to the minimum. Therefore, DDoS attack detection and defense still attract much concern from researchers.

II.METHODOLOGY

Sensor nodes are randomly distributed in the sensing field. In this network, the nodes are static and fixed. The sensor nodes sense the information and then send to the server. If the source node sends the packet, it will send through the intermediate node. The nodes are communicates only within the communication range. So, we have to find the node's communication range.

The cluster analysis consists of CURE algorithm which is efficient for large data scales. CURE is more robust to identify outliers and clusters with non-spherical shapes. It identifies the clustered data using a certain number of well-scattered points (the farthest from centroid), and then it uses the shrink factor to shift the scattered points toward the centroid by a specific fraction.

In addition, it is a point assignment algorithm that uses the Euclidean distance. This algorithm overcomes the traditional clustering algorithms that produce spherical shapes and similar sizes such as centroid-points assignment in that it handles not well-distributed data and identifies clusters with non-spherical shapes by introducing a new concept between the centroid and all the points in the cluster. In CURE, a collection of well-scatter points is used to identify the cluster shape.

III.DATA MINING ASSISTS IN INTRUSION DETECTION

The central theme of intrusion detection using data mining approach is to detect the security violations in information system. Data mining can process large amount of data and it discovers hidden and ignored information. To detect the intrusion, data mining consists of following processes such as classification, clustering, and regression [3]. It monitors the information system and raises alarms when security violations are founded.

Genetic Algorithms Genetic algorithms were initially introduced in the meadow of computational biology. After that they have been bloomed into various fields with promising result [24]. Nowadays the researchers have tried to incorporate this algorithm with IDSs. Using Genetic approach, in 1995 Giordana and Neri has proposed one intrusion detection algorithms called REGAl. The REGAL System is based on distributed genetic algorithm. REGAL is a concept learning system that learns First Order Logic multi-model concept descriptions. The learning examples are stored in relational database that are represented as relational tuples. Gonzalez and Dasgupta [26] applied a genetic algorithm, though they were examined host based IDSs, not network based. They used the algorithm only for the Meta learning step instead of running algorithm directly on the feature set. It uses the statistical classifiers for labelled vectors. A 2-bit binary encoding methodology is used for identifying the abnormality of a particular feature, ranging from normal to abnormal. Chittur [27] used a genetic algorithm with decision tree. Decision tree is used to represent the data. They used the high detection rate that reduces the false positive rate. The false positive occurrence was minimized by utilizing human input in a feedback loop [10].

Bayesian Classifier A Bayesian Classifier provides high accuracy and speed for handling large database. In network model Bayesian classifier encodes the probabilistic relationship among the variable of interest. In intrusion detection this classifier is combined with statistical schemes to produce higher encoding interdependencies between the variables and predicting events. The graphical model of casual relationships performs learning technique. This technique is defined by two components-a directed acyclic graph and a set of conditional probability tables. Direct Acyclic Graph (DAG) represents a random variable, which may be discrete or continuous. For each variable classifier maintain one conditional probability table (CPT) and it requires higher computational effort[12].

Proposed Algorithm
Step-1: Start the weka tool.
Step2: Open Browse the Data base.
Step3: Convert .csv file format to .arff format and process data.
Step4: Apply the feature selection method to select the feature.
Step5: Apply different classification methods to classify the data.
Step6: Calculate the different parameters in the form of TP, TN,FP and FN.
Step7: Stop.

IV. A COMPARATIVE ANALYSIS OF DATA MINING TECHNIQUES FOR INTRUSION DETECTION SYSTEM

Classifier	Method	Advantages	Disadvantages
Support Vector Machine	A support vector machine is a classification and regression technique it constructs a hyper plane or set of hyper planes in a high or infinite dimensional space.	1. High Accuracy. 2. Able to model complex and nonlinear decision boundaries. 3. Less prone to over fitting than other methods.	1. High algorithmic complexity and extensive memory requirement. 2. The choice of the kernel is difficult. 3. The training and testing speed is slow
Genetic Algorithm	Genetic algorithm learning examples are stored in relational database that are represented as relational tuples.	1. It solves every optimization problem. 2.It solves the problems with multiple solutions 3. Easily transferred to existing models.	1. No global optimum. 2. No constant optimization response time
K Nearest Neighbour	An object classification process is achieved by the majority vote of its neighbours. The object is being assigned to the class most common amongst its k nearest neighbours. If k = 1, then the object is simply assigned to the class of its nearby neighbour.	1. Analytically tractable. 2. Implementation task is simple. 3.Highly adaptive behaviour 4.Easy for parallel implementations	1. High storage requirements. 2. Highly susceptible to the curse of dimensionality. 3. Slow in classifying and testing tuples.
Neural Network	A Neural Network is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase.	1. Requires less formal statistical training. 2. Implicitly detect the complex nonlinear relationships between dependent and independent variables. 3. Highly tolerate the noisy data. 4. Availability of multiple training algorithms.	1. Process is black box. 2. Greater computational burden. 3. Over fitting. 4. It Requires long training time.
Bayesian Method	Bayesian classifier based on	1. Naïve Bayesian classifier	1. The assumptions made in

	the rules. It uses the joint probabilities of sample classes and observations. The algorithm tries to estimate the conditional probabilities of classes given an observation.	simplifies the computations. 2. Exhibit high accuracy and speed when applied to large databases.	class conditional independence. 2.Lack of available probability data
Decision Tree	Decision tree initially builds a tree with classification. Each node represents a binary predicate on one attribute, one branch represents the positive instances of the predicate and the other branch represents the negative instances.	1. Construction does not require any domain knowledge. 2. Can handle high dimensional data. 3. Representation is easy to understand. 4. Able to process both numerical and categorical data.	1. Output attribute must be categorical. 2. Limited to one output attribute. 3. Decision tree algorithms are unstable. 4. Trees created from numeric datasets can be complex.
Fuzzy Logic	The fuzzy logic has been used for both anomaly and misuse intrusion detection.	1. Uses linguistic variables. 2. Allows imprecise inputs. 3.Permits fuzzy thresholds 4.Reconciles conflicting objectives 5.Rule base or fuzzy sets easily modified	1. Hard to develop a model from a fuzzy system. 2. Require more fine tuning and simulation before operational.

V.RESULT&DISCUSSION

In this research work intrusion is detected with the help of weka tool and Matlab. Here Weka tool Snap Shorts is given below:

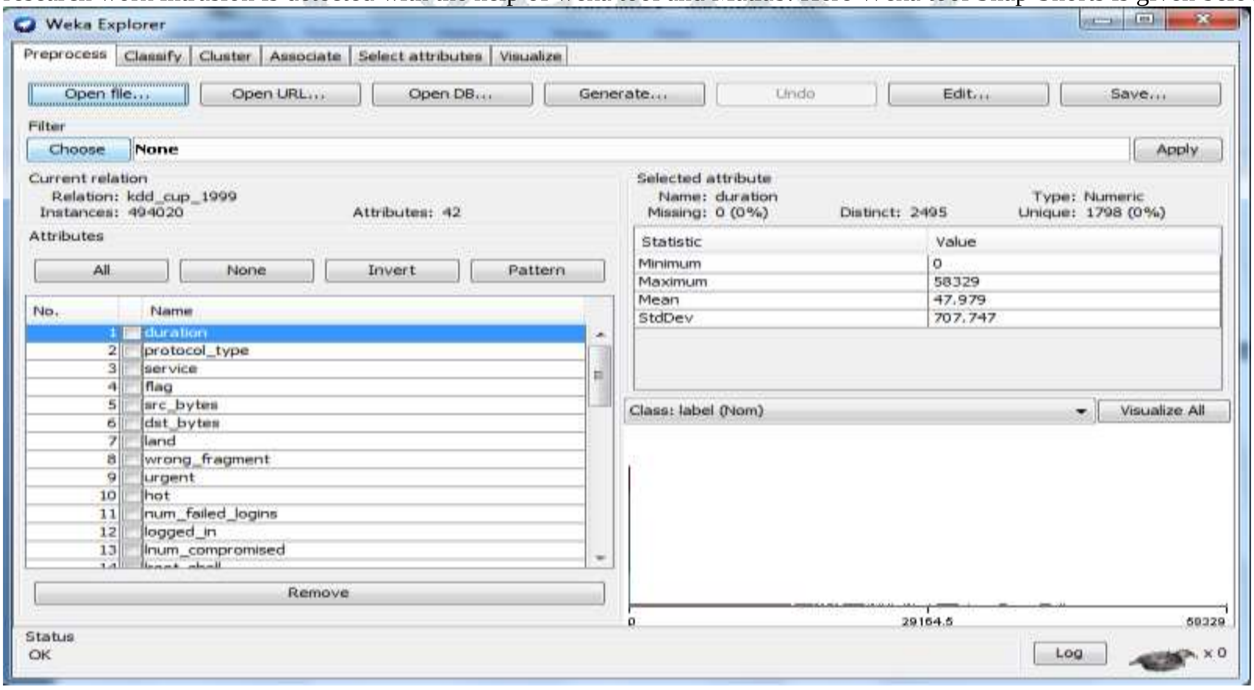

Figure 1: KDDcups network database

The figure 1 is defining the KDDcups network database. It is used to display the different features of the database. It is also defining the minimum, maximum, Mean and StdDev Values.

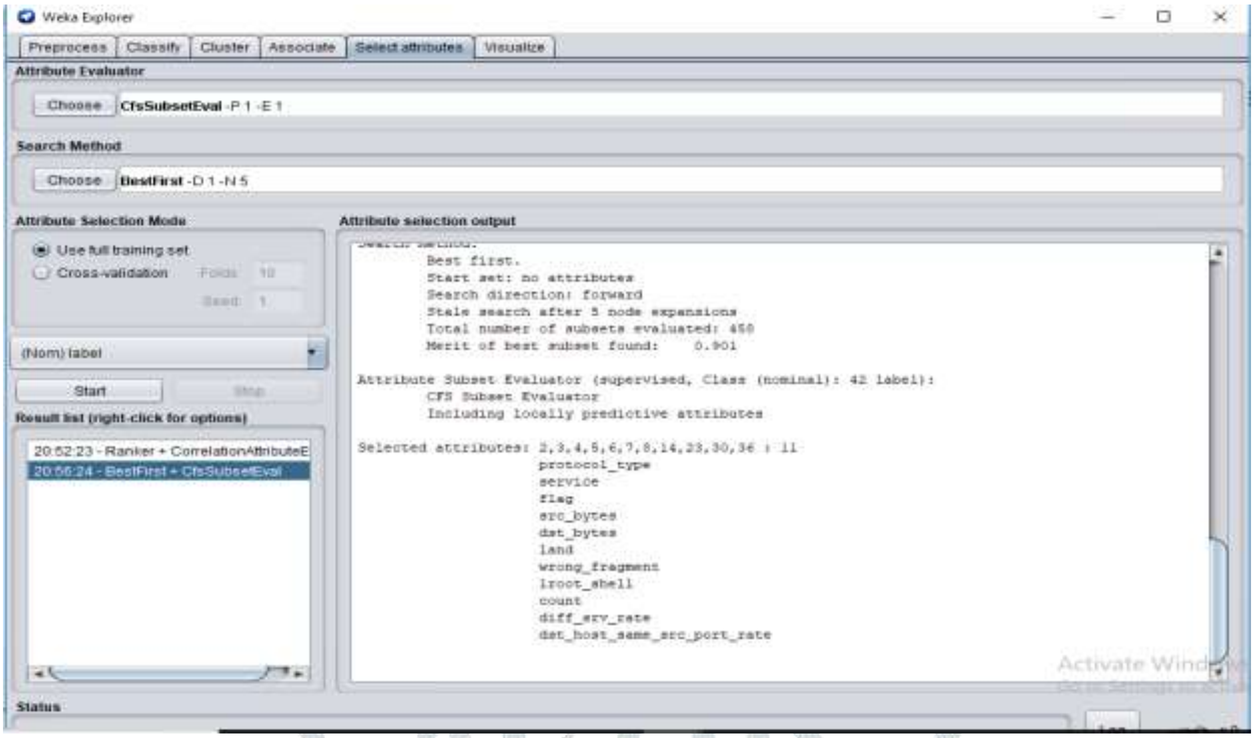

Figure 2:Selected attributes: 2,3,4,5,6,7,8,14,23,30,36 : 11

The figure 2 is defining the selected attributes of the database. In this figure different attributes like 2,3,4,5,6,7,8,14,23,30,36 : 11 are selected for performing the different operations.

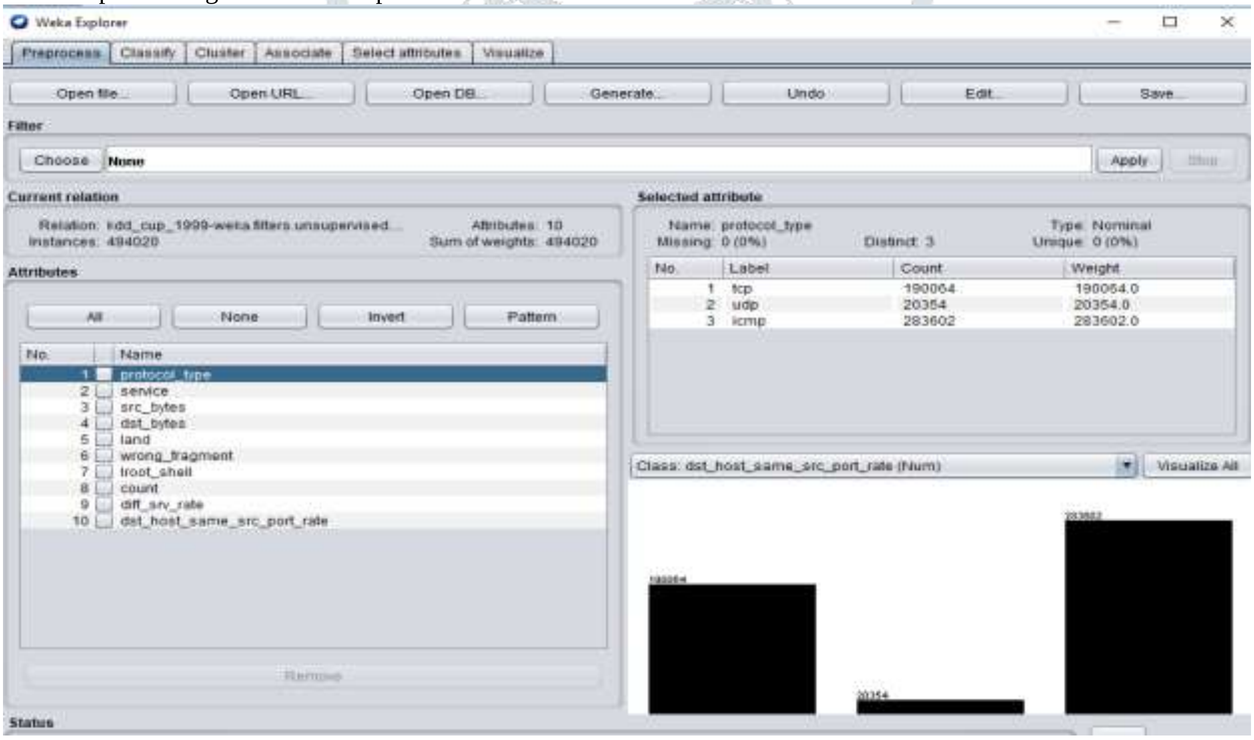

Figure 3: After Feature selection

The figure 3 is displaying the selected features after the figure 5.4. because in the figure 5.4 only features are selected and in this figure non selected features are removed and only selected features are kept for processing.

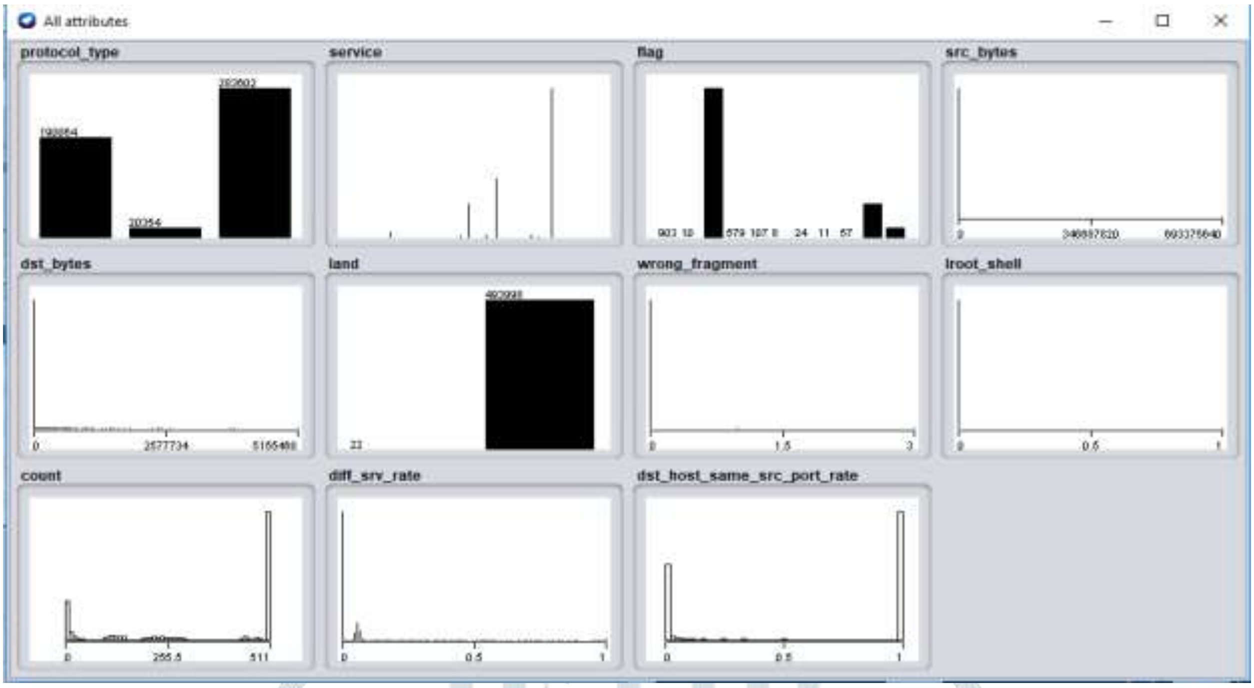

Figure 4: Visualization of classified attributes

The figure 4 is the graphical visualization of the classified attributes after processing. This figure shows only the selected features in the form of Graphs.

KStar options

Time taken to build model: 0.03 seconds

Table 1: Stratified cross-validation KStar options

Correctly Classified Instances	10789	91.0464 %
Incorrectly Classified Instances	1061	8.9536 %

Table 2: Performance parameters using KStar

Kappa statistic	0.6688
Mean absolute error	0.157
Root mean squared error	0.2726
Relative absolute error	52.8008 %
Root relative squared error	70.7073 %
Coverage of cases (0.95 level)	99.7046 %
Mean rel. region size (0.95 level)	95.1519 %
Total Number of Instances	11850

Table 3: Detailed Accuracy by KStar

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.637	0.029	0.830	0.637	0.721	0.677	0.880	0.751	normal
0.971	0.363	0.923	0.971	0.947	0.677	0.880	0.953	anomaly

Table 4: Confusion Matrix KStar

a	b	classified as
1371	781	a = normal
280	9418	b = anomaly

Multiclass Classifier

Time taken to build model: 0.91 seconds

Table 5: Stratified cross-validation Multiclass Classifier

Correctly Classified Instances	10266	86.6329 %
Incorrectly Classified Instances	1584	13.3671 %

Table 6: Performance parameters using Multiclass Classifier

Kappa statistic	0.5584
Mean absolute error	0.2039
Root mean squared error	0.3195
Relative absolute error	68.5803 %
Root relative squared error	82.8854 %
Coverage of cases (0.95 level)	99.9831 %

Mean rel. region size (0.95 level)	99.9578 %
Total Number of Instances	11850

Table 7: Detailed Accuracy by Multiclass Classifier

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.656	0.087	0.626	0.656	0.640	0.559	0.817	0.536	normal
0.913	0.344	0.923	0.913	0.918	0.559	0.817	0.929	anomaly

Table 8: Confusion Matrix Multiclass Classifier

a	b	classified as
1411	741	a = normal
843	8855	b = anomaly

DARPA DATASET

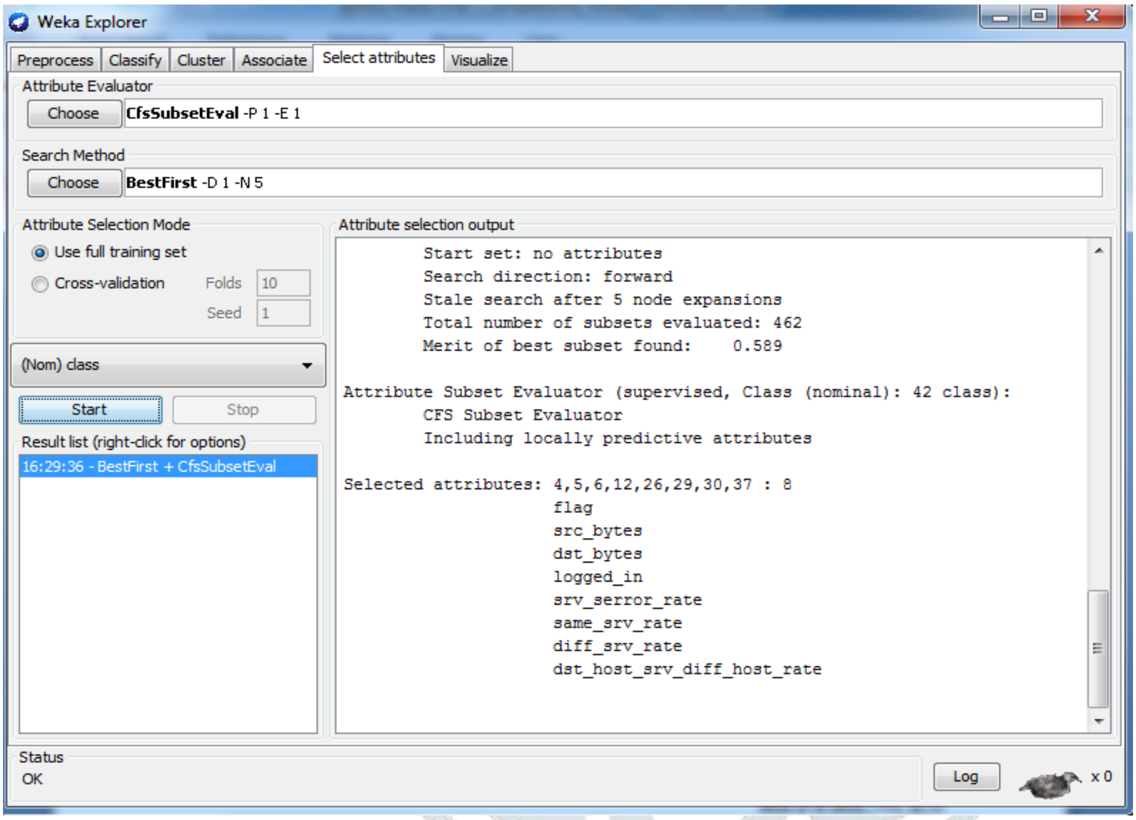

Figure 5: DARPA dataset selected attributes: 4, 5, 6,12,26,29,30,37: 8

Time taken to build model: 19.32 seconds

Table 8: Stratified cross-validation

Correctly Classified Instances	24504	97.269 %
Incorrectly Classified Instances	688	2.731 %

Table 9: Performance parameters on DARPA dataset

Kappa statistic	0.9451
Mean absolute error	0.0403
Root mean squared error	0.1438
Relative absolute error	8.1051 %
Root relative squared error	28.8333 %
Coverage of cases (0.95 level)	99.5514 %
Mean rel. region size (0.95 level)	55.4779 %
Total Number of Instances	25192

Table 10: Detailed Accuracy on DARPA dataset

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.979	0.034	0.970	0.979	0.975	0.945	0.994	0.990	normal
0.966	0.021	0.975	0.966	0.971	0.945	0.994	0.994	anomaly

Table 11: Confusion Matrix

a	b	classified as
13164	285	a = normal
403	11340	b = anomaly

VI.CONCLUSION&FUTURE WORK

Distributed Denial of Service attacks (DDoS) overwhelm network resources with useless or harmful packets and prevent normal users from accessing these network resources. These attacks jeopardize the confidentiality, privacy and integrity of information on the internet Network security is one of the most important issues that can be considered by commercial organizations to protect its information from malicious risk. The problems of detection malicious traffics have been widely studied and still as a hot research topic in the recent decades. Sensor nodes are randomly distributed in the sensing field. In this network, the nodes are static and fixed. The sensor nodes sense the information and then send to the server. If the source node sends the packet, it will send through the intermediate node. The nodes are communicates only within the communication range. So, we have to find the node's communication range. In this work different problems are resolved and different parameters like TP,FP, Precision and Recall is calculated. Here 94% result is achieved in DARPA dataset with fuzzy logic. In future it is extended on different datasets with different classifiers.

REFERENCES

[1] Review of DDOS Attack and its Countermeasures in TCP Based Networks" International Journal of Computer Science & Engineering Survey (IJCSES) Vol.2, No.4, November 2011.

[2] Divya Kuriakose,V.Praveena "A Survey on DDoS Attacks and Defense Approaches" International Journal of Innovative Research in Computer and Communication Engineering ,Vol. 1, Issue 8, October 2013.

[3] Saurabh Ratnaparkhi , Anup Bhange " Protecting Against Distributed Denial of Service Attacks and its Classification: An Network Security Issue" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 1, January 2013

[4] Darshan Lal Meena, Dr. R. S. Jadon " Distributed Denial of Service Attacks and Their Suggested Defense Remedial Approaches" International Journal of Advance Research in Computer Science and Management Studies , Volume 2, Issue 4, April 2014.

[5] Shenam Chugh, Dr. Kamal Dhanda " Denial of Service Attacks" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 8, August 2015

[6] Raksha Upadhyay, Uma Rathore Bhatt, Harendra Tripathi "DDOS Attack Aware DSR Routing Protocol in WSN" International Conference on Information Security & Privacy (ICISP2015), 11-12 December 2015.

[7] Wesam Bhaya, Mehdi EbadyManaa "DDoS Attack Detection Approach using an Efficient Cluster Analysis in Large Data Scale" Annual Conference on New Trends in Information & Communications Technology Applications-(NTICT'2017) 7 - 9 March 2017.

[8] Liang Hu, Xiaoming Bi, "Research of DDoS Attack Mechanism and Its Defense Frame,"Computer Research and Development (ICCRD), 3rd International Conference, pp. 440–442, March 2011.

[9] Robert Vamosi, "Study: DDoS attacks threaten ISP infrastructure," Online at http://news.cnet.com/8301-1009_3-10093699-83.html, CNET News, Nov. 2008.

[10] Elinor Mills, "Radio Free Europe DDOS attack latest by hactivists," Online at http://news.cnet.com/8301-10784_3-9933746-7.html, CNET News, May. 2008.

[11] Christos Douligeris and Aikaterini Mitrokotsa, "DDoS Attacks And Defence mechanisms: A Classification," in Proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology, (ISSPIT'03), pp. 190-193, Dec 2003.

[12] Nisha H. Bhandari, "Survey on DDoS attacks and its detection defense approach," International Journal of Science and Modern Engineering,Vol.1, Issue.3, pp.67-71, Feb 2013.

[13] S.A.Arunmozhi, Y.Venkataramani,"DDoS attack and Defense in wireless ad-hoc Network," International Journal of Network Security & Its Applications Vol.3, No.3, pp.182-187, May 2011.

[14] Monika Sachdeva, Gurvinddr Singh, Krishnan Kumar, Kuldip Singh, " A comprehensive Survey of Distributed Defense Techniques against DDoS Attack," International Journal of Computer Science and Network Security, Vol.9, No.12, pp.7-15, Dec 2009.

[15] Shibiao Lin Tzi-cker Chiueh," A Survey on Solutions to Distributed Denial of Service Attacks", Department of Computer Science Stony Brook University, pp.1-38, Sep 2006.

[16] Shuchi Juyali, Radhika Prabhakar, "A Comprehensive Study of DDOS Attacks and Defense Mechanisms," Journal of Information and Operations Management, Vol.3, Issue.1, 2012.

[17] Quan Jia, Kun Sun, Angelos Stavrou, "CapMan: Capability-based Defense against Multi-Path Denial of Service (DoS) Attacks in MANET,"proceedings of the 20th international conference on computer communication and networks, pp 1-6, 2011.

[18] Antonio Challita, Mona El Hassan, Sabine Maalouf, Adel Zouheiry, " A Survey of DDoS Defense Mechanisms ," The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[19] Anurekha, R.,K. Duraiswamy, A. Viswanathan, V.P. Arunachalam, K. Ganesh Kumar, A. Rajivkannan" Dynamic Approach to Defend Against Distributed Denial of Service Attacks Using an Adaptive Spin Lock Rate Control Mechanism," Journal of Computer Science, pp.632-636, 2012.

[20] Puneet Zaroo," A Survey of DDoS attacks and some DDoS defense mechanisms," Advanced Information Assurance (CS 626), 2003.

[21] Guangsen Zhang, Manish Parashar,"Cooperative Defense against DDoS Attacks," Journal of Research and Practice in Information Technology, pp.1-6, 2006.

[22] Wei Ren, Dit-Yan Yeung, Hai Jin, Mei Yang, "Pulsing RoQ DDoS Attack and Defense Scheme in Mobile Ad Hoc Networks," International Journal of Network Security, Vol.4, No.2, pp.227-234, Mar. 2007.