DOI: 10.21917/ijct.2017.0224

PREVENTION OF ZERO DAY VULNERABILITY IN NETWORK USING ENSEMBLE FUZZY ASSOCIATION AND CUTTLE FISH DETECTION

M. Masthan[1] and R. Ravi[2]

[1]*Department of Computer Science and Engineering, Manonmaniam Sundaranar University, India*
[2]*Department of Computer Science and Engineering, Francis Xavier Engineering College, India*

Abstract

The data communication between different parts of the universe is managed by the computer networks and the Enterprise Information System (EIS) which rely on them. The privacy and security are the most important factor to be maintained in any network systems. This paper deals about the detection of intrusion attack in the eclipse database using Ensemble fuzzy association (EFA) and Cuttle Fish Algorithm (CFA). The proposed methodology creates a rule-based ensemble model for network diversity metric modeling for the efficient detection of zeroday attacks and to reduce the time consumption. The simulation result shows that the EFA and CFA having efficient detection rates as compared to the existing systems.

Keywords:

Enterprise Information System (EIS), Ensemble Fuzzy Association (EFA), Cuttle Fish Algorithm (CFA)

1. INTRODUCTION

Nowadays, the exchanges of large size information between the peoples situated around the world are governed by computer networks and they act as the backbone of Enterprise Information System (EIS). The large size of enterprises available in real-world employs the EIS to communicate with the customers [1], [2], and [4]. The extensive utilization of computer networks with critical infrastructures such as military organization, protecting mission, demands the placing the vulnerabilities [5] and the deployment of firewalls. According to the recent evidence report such as modern malware exploits the multiple unknown vulnerabilities, equal treatment of unknown vulnerabilities is necessary to improve the resilience of network against the attacks. The dealing of unknown vulnerabilities is the challenging task and it is regarded as the zero-day attacks and the lack of information [6] introduces the difficulties in mitigation. The evolution of target defense and diversity strategies in research studies rely on intuitive notions of diversity. Hence, the impact of network diversity on the security has received the limited attention [6] which is considered as the background of the proposed research. Our proposed research deals with the detection of intrusion attack in the eclipse database with the help of Ensemble Fuzzy Association (EFA) and Cuttle Fish Algorithm (CFA). The data communication between different parts of the universe is accomplished by the computer networks and the Enterprise Information System (EIS) which rely on them. The privacy and security are the most important factor to be maintained in any network systems. This methodology creates a rule-based ensemble model for network diversity metric modelling for the efficient detection of zero-day attacks and to reduce the time consumption. The proposed EFA and CFA have efficient detection rates when compared to the other existing systems.

The results obtained from the simulation shows that, the proposed method is capable of detecting the zero-day attack, when compared to the existing methodologies. The time consumption in detection of zero-day attack affects the accuracy imprecision over the approaches. The parameters like execution time, CPU utilization, false positive rate and attack detection rate are used to determine the performance analysis of proposed methodology. The optimistic results were observed which signifies the importance of proposed work. Thus the proposed Cuttle-Fish Detection and the Ensemble Fuzzy Approach provides better efficiency in the zero-day attack detection. Further extension of this research will be focusing on the efficient prevention of the zero-day attack in the eclipse platform.

2. RELATED WORK

The improvement of the resilience of software and networks depends on the security mechanism called network diversity [4]. The existing methods relied on the intuitive and imprecise notions of diversity and they are designed for single system software replicas or variants [1] [3]. During the higher abstraction stage, the network diversity has the great impact on the security and it requires limited attention. The existing method focuses on the modeling of network diversity as the security metric by using mathematical models correspond to the biodiversity of Ecology. Initially, the metric counts the number of distinct resources inside the network which facilitate the equal treatment of all the uneven resources. Second, the network diversity metric is designed [5] based on the attacking effort and their effect on the relationship between the resources. Third, the complementary metric called probabilistic network diversity security metric is modeled [6] to reflect the average attacking effort. Finally, the highlighted three metrics are validated with the various simulations under different cases. The ignorance of potential casual relationships between the resources affects the accuracy of detection of the zero-day attack.

3. PROPOSED TECHNIQUE

To overcome the limitations, we propose the combination of two algorithms such as Ensemble Fuzzy Association (EFA) and Cuttle Fish Detection (CFD) explained in Fig.1. The input dataset is split up into several subsets through the preprocessing stage which is considered as the initial process in proposed work.

Based on the metrics such as support and confidence, the association rules [7] are generated with the ensemble process to predict the severity clusters. These clusters are considered as the initial population for CFD algorithm. By using the fitness value, the attackers that affect the network diversity metric are detected [5] and isolated from the normal. The employment of parallel

processing through the EFA and CFD models reduces the computational time of the network effectively with increased throughput.

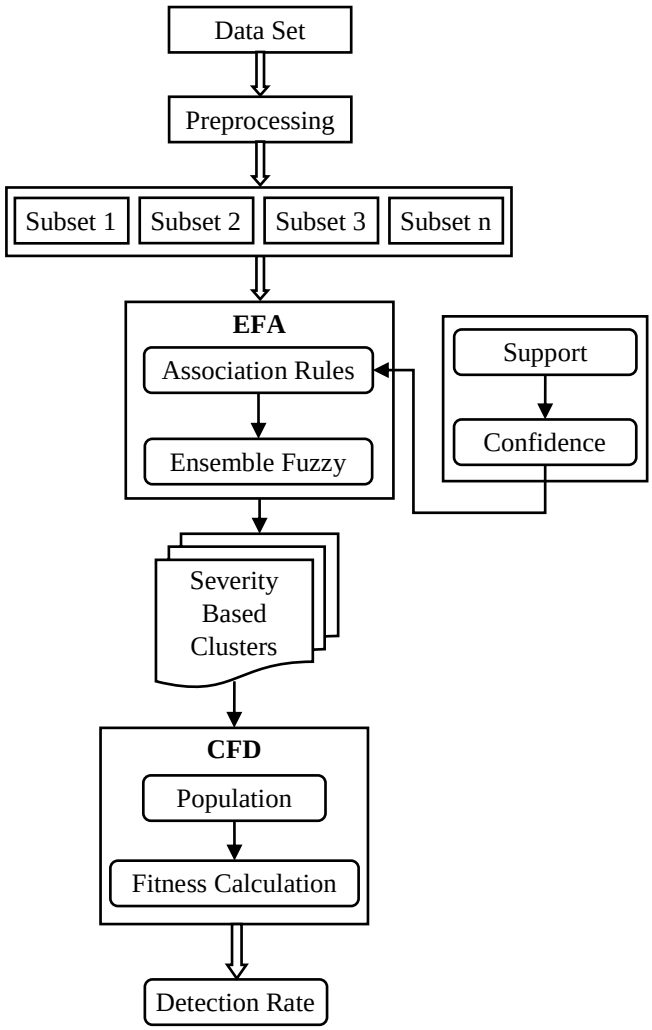

Fig.1. Proposed Architecture

3.1 WORKING MODULE

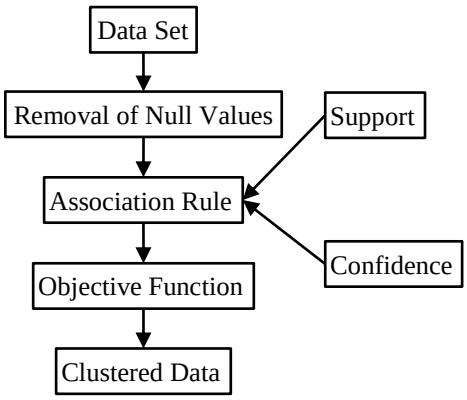

Fig.2. Ensemble Fuzzy Approach (EFA)

The working principle of the proposed methodology comprises of two stages, one is the clustering stage accompanied by the EFA and the other one is the detection stage carried out by the Cuttle Fish Detection (CFD). The following Fig.2 and Fig.3 shows the Flow chart containing the working mechanism of the suggested method.

Ensemble fuzzy association is used to create clusters. It includes association rules. Association rules are used to find the relationship between the attributes. For creating association rules support and confidence is used. Support is calculated between the items in database. It is used to find the frequency of the items in the database. The Eq.(1) is used to calculate support:

$$Su = \frac{\sum_{1}^{n} Occ}{\sum_{1}^{n} To} \tag{1}$$

Confidence is used to find how often the rule has been found to be true. Based on the two support values confidence can be calculated. The Eq.(2) is used to calculate confidence:

$$Co = \frac{Su_{jk}}{Su_{j}} \tag{2}$$

From the confidence, threshold is calculated to filter the formed association rules. The Eq.(3) is used to calculate threshold:

$$Th = \frac{\sum_{1}^{n} Co}{Co.Size} \tag{3}$$

The filtered data is given as input to clustering. For create cluster, user need to mention number of clusters to be formed. Based on that cluster heads are selected. Each data is compared with each cluster head, and calculate objective function. Based on the calculated objective function clusters are formed. The Eq.(4) - Eq.(6) are used to calculate objective function:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{ik}} \right)^{\left(\frac{2}{m-1} \right)}} \tag{4}$$

$$v_j = \frac{\sum_{i=1}^{n} \left(\mu_{ij} \right)^m x_i}{\sum_{i=1}^{n} \left(\mu_{ij} \right)^m} \tag{5}$$

$$J\left(U,V\right) = \sum_{i=1}^{n} \sum_{j=1}^{c} \left(\mu_{ij} \right)^m \left\| x_i - v_j \right\|^2 \tag{6}$$

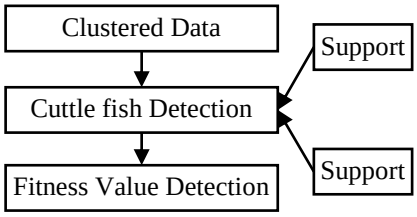

Fig.3. Cuttle Fish Detection (CFD)

Cuttle fish detection is used to find the vulnerability in dataset. These vulnerabilities are mainly focused by the attackers. Before that, developers need to find the vulnerabilities in dataset. It is used to avoid the attack executed by attackers. For Cuttle Fish Detection clustered data is given as input. It is called as

population. Based on the population condition and action is calculated. Condition and action is used to find the fitness values. If confidence and cluster head are equal means condition is increased. If confidence alone is equal means action will be increased. AB is condition. A is action. The Eq.(7) is used to find the fitness:

If $Con_i == Con_j$ && $Ac_i == Ac_j$

AB++

else if $Con_i == Con_j$

A++

$$F_i = 2 + \frac{AB - A}{AB + A} + \frac{AB}{X} - \frac{A}{Y} \quad (7)$$

3.2 ENSEMBLE FUZZY APPROACH

The efficient detection of intrusion involves the classification o subsets followed by the Association rules [7]. The Associations rules are incorporated to provide a normalized cluster of information matrices for the input to the ensemble fuzzy block. The following algorithm will enumerate the proposed technique.

Algorithm 1: Ensemble Fuzzy Approach (EFA)

Initialization of Dataset

ID ← load dataset

for i=0: ID do

D ← Filter Valid Data

end for

Nos ←no of subsets

for i=0: Nos do

Is ← initialize subsets

end for

Procedure:

A ← attributes of dataset

Su ← support

Occ ← occurrence

To ← total

for i=0: Nos do

for j=0: A do

for k=0: A do

$$Su = \frac{\sum_{1}^{n} Occ}{\sum_{1}^{n} To}$$

end for

end for

end for

Co ← Confidence

for i=0: Nos do

for j=0: Att do

for k=0: Att do

$$Co = \frac{Su_{jk}}{Su_j}$$

end for

end for

end for

Th ← threshold

$$Th = \frac{\sum_{1}^{n} Co}{Co.Size}$$

Noch ← no of cluster head

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{ik}} \right)^{\left(\frac{2}{m-1} \right)}}$$

$$v_j = \frac{\sum_{i=1}^{n} \left(\mu_{ij} \right)^m x_i}{\sum_{i=1}^{n} \left(\mu_{ij} \right)^m}$$

Of ← objective function

$$Of \leftarrow J(U,V) = \sum_{i=1}^{n} \sum_{j=1}^{c} \left(\mu_{ij} \right)^m \left\| x_i - v_j \right\|^2$$

Output:

Data set into pure clustered dat.

3.3 CUTTLE FISH DETECTION

Algorithm 2: Cuttle Fish Detection (CFD)

Po ← population

Po ← clustered data

Con ← condition

Con ← Co

Ac ← action

Ac ← allocated clusters

Fi ← fitness

Procedure:

for i = 0: Po do

for j = 0: Po do

if $Con_i == Con_j$ && $Ac_i == Ac_j$

AB++

else

A++

end for

end for

X = maximum of AB in Po

Y = maximum of A in Po

$$F_i = 2 + \frac{AB - A}{AB + A} + \frac{AB}{X} - \frac{A}{Y}$$

Ref ← reflection

Vis ← visibility

Subset ← *Ref* + *Vis*

Output:

Detected data set of vulnerable and non-vulnerable data sets.

The data sets finally will be clustered and formed as sub divided groups. Thus, the clustering of the sensor nodes are manipulated by the Cuttle Fish algorithm. And the later stage of intrusion detection phase is obtained by the Ensemble fuzzy algorithm.

4. SIMULATION RESULTS

This section provides the simulation and analysis of the defined Cuttle Fish Detection using ensemble fuzzy approach.

The Fig.4 shows the loading of data sets having attributes such as No. of Bugs found, No. of Non-trivial Bugs Found, No. of Major Bugs Found, No. of Critical Bugs etc.

The Fig.5 shows that the selection of subsets, i.e. dividing of the overall classes into multiple subsets to clean the unwanted data sets.

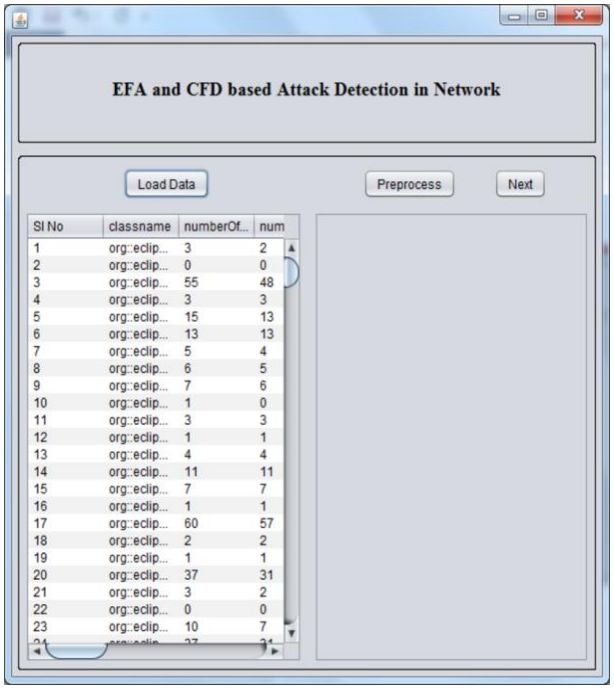

Fig.4. Loading data sets

The Fig.6 depicts that the selection of subsets for the manipulation of threshold value to normalize the data sets in order to process the association rule.

The Fig.7 reveals about the functions called support and confidence, which compares and identifies the relationship between the columns of the data sets or classes.

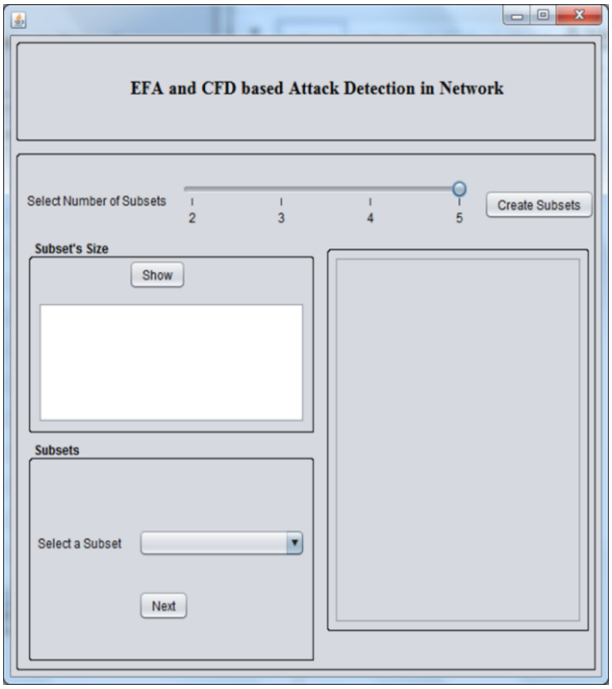

Fig.5. Selection of No. of Sets

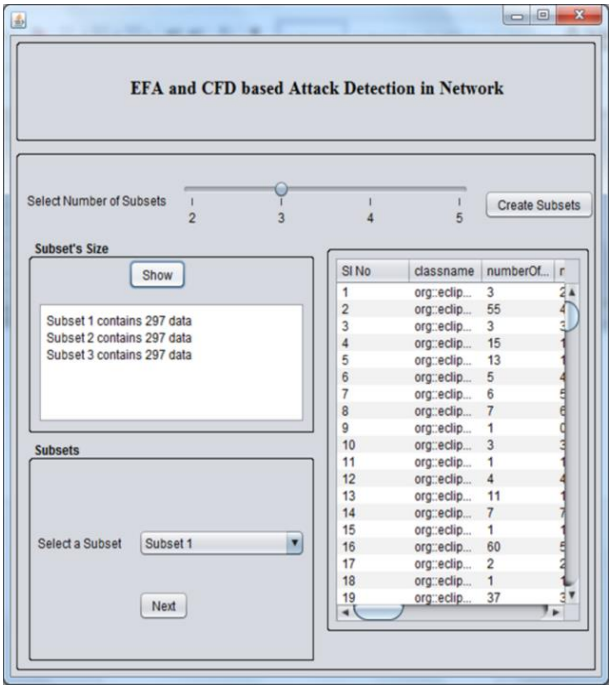

Fig.6. Selection of Subset

The Fig.8 shows that the identification of threshold values among the clusters, and to obtain the normalized value among the cluster heads.

The Fig.9 depicts that the formation of objective fitness obtained from the sequential processing of fuzzy clustering followed by CFD.

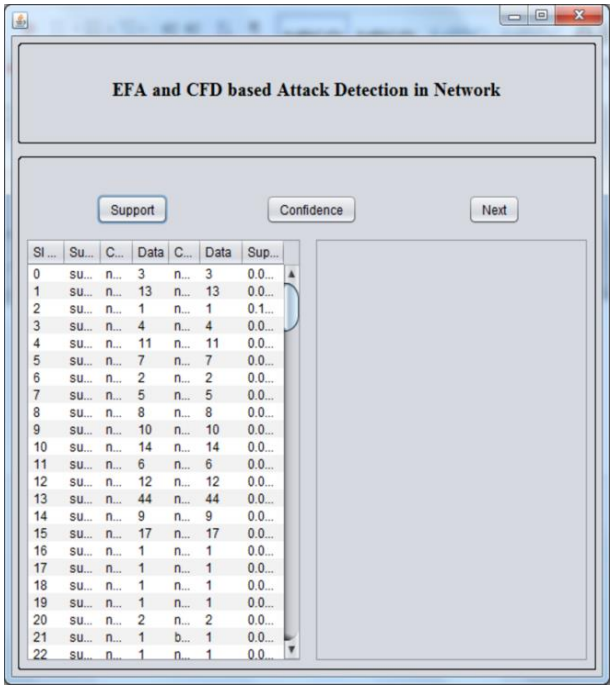

Fig.7. Association Rule Formulation

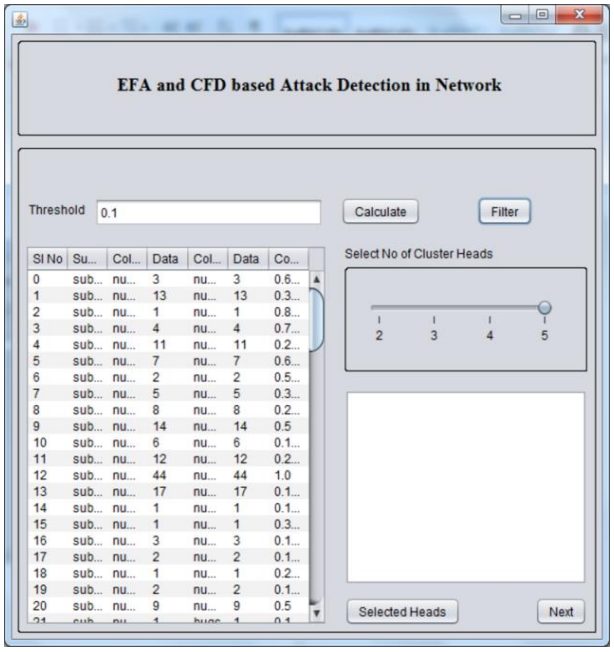

Fig.8. Threshold Manipulation

5. PARAMETRIC ANALYSIS

The parameters that are analyzed in this paper are Execution time, CPU Utilization, False positive rate and Attack Detection Rate. The graphical representation of the obtained simulation results were shown below.

5.1 EXECUTION TIME

The Execution Time is the run time duration in which a program is running (executing) in contrast to other program life cycle phases.

$$\text{Execution Time} = CPI \times I \times C \quad (8)$$

where, *CPI* - Cycles per Instruction, *I* - Instruction and *C* - CPU Clock Cycle.

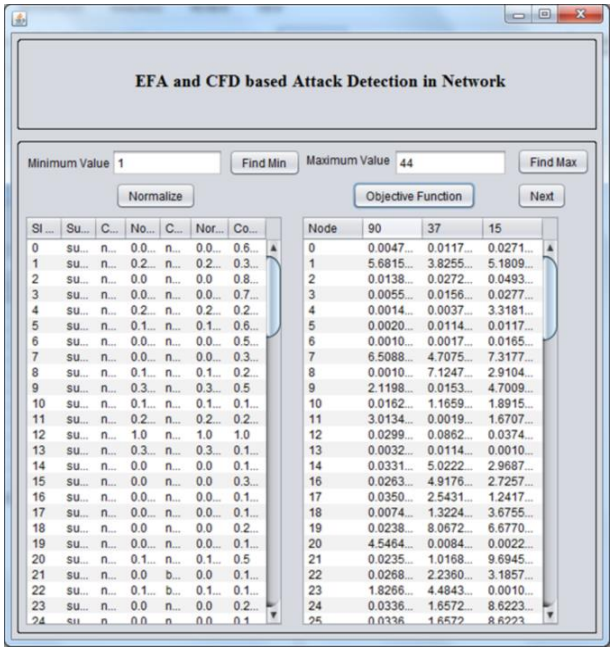

Fig.9. Objective Fitness Detection

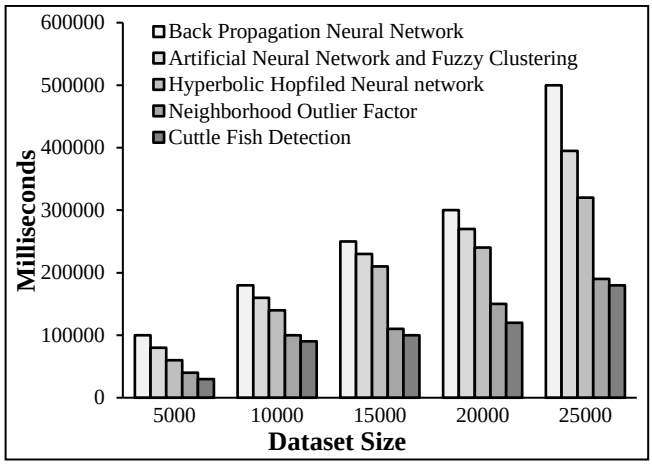

Fig.10. Execution Time vs. Dataset Size

From the Fig.10 we can clearly say that the execution time of our proposed Cuttle Fish Detection (CFD) outperforms among the existing techniques such as back propagation neural network, Artificial Neural Network and Fuzzy clustering, Neighborhood outlier factor.

5.2 CPU UTILIZATION

The CPU utilization is the computer usage of processing the resources, or the amount of work handled by the CPU.

$$\textit{CPU Utilization} = 100 - (\%\ \textit{of time spent in idle task}) \quad (9)$$

From the Fig.11 it is clear that, the CPU Utilization time is very less in our proposed CFD technique than to the other existing methods.

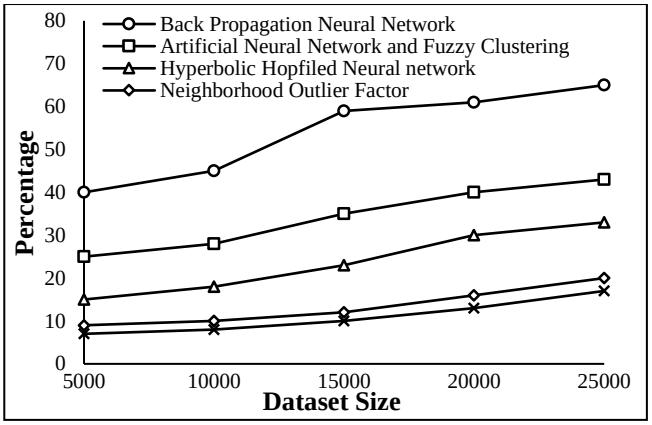

Fig.11. CPU Utilization vs. Data Size

5.3 FALSE POSITIVE RATE

The False Positive Rate also called as the False Alarm Ratio can be defined as the probability of falsely rejecting the null hypothesis for a particular set.

$$FPR = FP / N = FP / FP + TN \quad (10)$$

where, *FP* - No of false positives, *TN* - True Negatives and *N* - Total No. of Negatives.

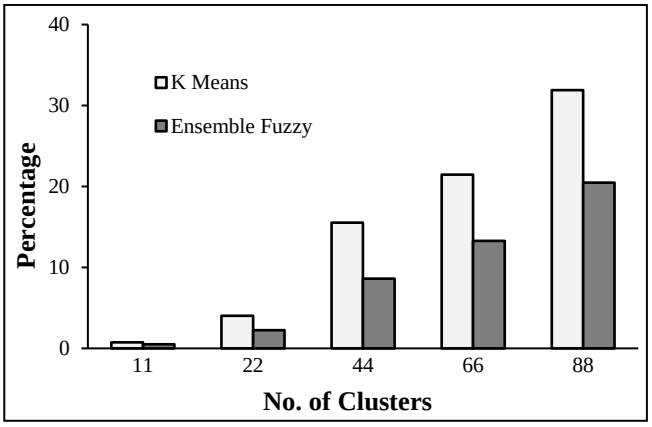

Fig.12. False Positive Rate vs. No of Clusters

Here we are comparing the FPR for the K-Means Clustering and our proposed Ensemble Fuzzy technique, and it is clear from the Fig.12 that, our proposed *EFA* having lesser *FPR* when compared with the K-Means clustering.

5.4 ATTACK DETECTION RATE

The Number of successful detection of intrusion or attacks in the executing platform are called as the Attack Detection Rate. Here in the graphical representation shown in the Fig.13 we can depict the fact that, our proposed Ensemble Fuzzy technique detects the intrusion in a higher rate than in comparison with the K-Means Clustering.

6. CONCLUSION

The results obtained from the simulation shows that, the proposed method is capable of detection the zeroday attack, when compared to the existing methodologies. As the time consumption in detection of the zeroday attack and the accuracy imprecision over the other approaches makes the suggested ensemble fuzzy a lead. The analysis has been made for the parameters such as Execution time, CPU utilization, False Positive rate and Attack detection rate. An optimistic results were observed as we can see in the graphical representation. Thus the proposed Cuttle Fish detection and the Ensemble fussy approach provides better efficiency in the zeroday attack detection. Further extension of this research will be focusing on the efficient prevention of the zeroday attack in the eclipse platform.

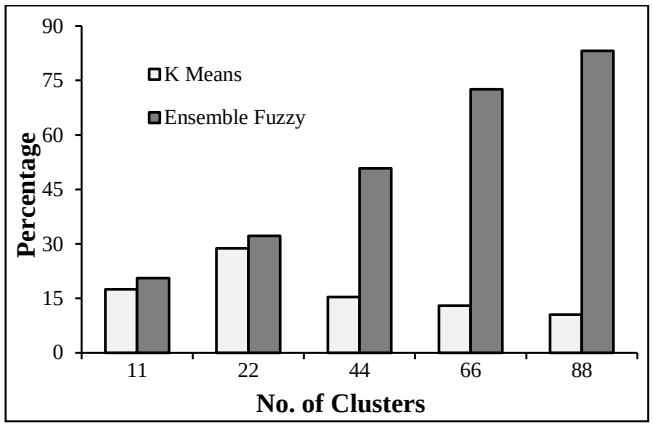

Fig.13. Attack Detection Rate vs. No. of Clusters

REFERENCES

[1] Nicolas Falliere, Liam O Murchu and Eric Chien, "W32.Stuxnet Dossier", Available at: https://scadahacker.com/library/Documents/Cyber_Events/Symantec%20-%20Stuxnet%20Dossier%20v1.4.pdf.

[2] Bev Littlewood and Lorenzo Strigini, "Redundancy and Diversity in Security", *Proceedings of European Symposium on Research in Computer Security*, pp. 423-438, 2004.

[3] Benjamin Cox, David Evans, Adrian Filipi, Jonathan Rowanhill, Wei Hu, Jack Davidson, John Knight, Anh Nguyen-Tuong and Jason Hiser, "N-Variant Systems: A Secretless Framework for Security through Diversity", *Proceedings of 15th USENIX Security Symposium*, pp. 105-120, 2006.

[4] Debin Gao, Michael K. Reiter and Dawn Song, "Behavioral Distance Measurement using Hidden Markov Models", *Proceedings of International Workshop on Recent Advances in Intrusion Detection*, pp. 19-40, 2006.

[5] Byung-Gon Chun, Petros Maniatis and Scott Shenker, "Diverse Replication for Single Machine Byzantine-Fault Tolerance", *Proceedings of USENIX Annual Technical Conference*, pp. 287-292, 2008.

[6] M. Garcia, A. Bessani, I. Gashi, N. Neves and R. Obelheiro, "OS Diversity for Intrusion Tolerance: Myth or Reality?", *Proceedings of IEEE/IFIP 41st International Conference on Dependable Systems and Networks*, pp. 383-394, 2011.

[7] Sandeep Bhatkar, Daniel C. DuVarney and R. Sekar, "Address Obfuscation: An Efficient Approach to Combat A Broad Range of Memory Error Exploits", *Proceedings of* 12th *USENIX Annual Technical Conference*, pp. 105-120, 2003.