INTRUSION DETECTION SYSTEM USING HIERARCHICAL CLUSTERING FROM DATASET

[1]Shikha Attri and [2]R C Gangwar And [3] Rajeev Bedi

[1]Post-Graduate Student, Computer Sc. & Engg, Beant College of Engg. & Tech, Gurdaspur (Pb) India

[2]Associate Professor, Department of Computer Sc, Beant College of Engg. & Tech, Gurdaspur(Pb) India

[3] Assistant Professor, Department of Computer Sc, Beant College of Engg. & Tech, Gurdaspur(Pb) India

ABSTRACT : The intrusion detection is the mechanism by which abnormality from the state driven dataset is discovered. The intrusion causes the problem of false discovery that mislead overall result. The resources from server may not be accessed by the use of intrusion be malicious users. The propose mechanism of hierarchical clustering to discover abnormal patterns from the dataset. The dataset is derived from kaggle website. The operation is demonstrated against K means clustering. The result is presented in terms of classification accuracy, false positive rate and false negative rate. the result shows significant improvement by the margin of 10%.

Keywords: Intrusion detection, false positive rate, classification accuracy, false negative rate

I. INTRODUCTION

The intrusion within the dataset causes significant affect on the classification accuracy. [1]The pattern discovery is hampered by the presence of intrusion within the dataset. The detection of abnormality within dataset is researched over by the use of clustering mechanism. [2]The clustering mechanisms that are commonly employed along with advantages and disadvantages are discussed in this literature. In addition attacks that are common on dataset are also elaborated through this literature.

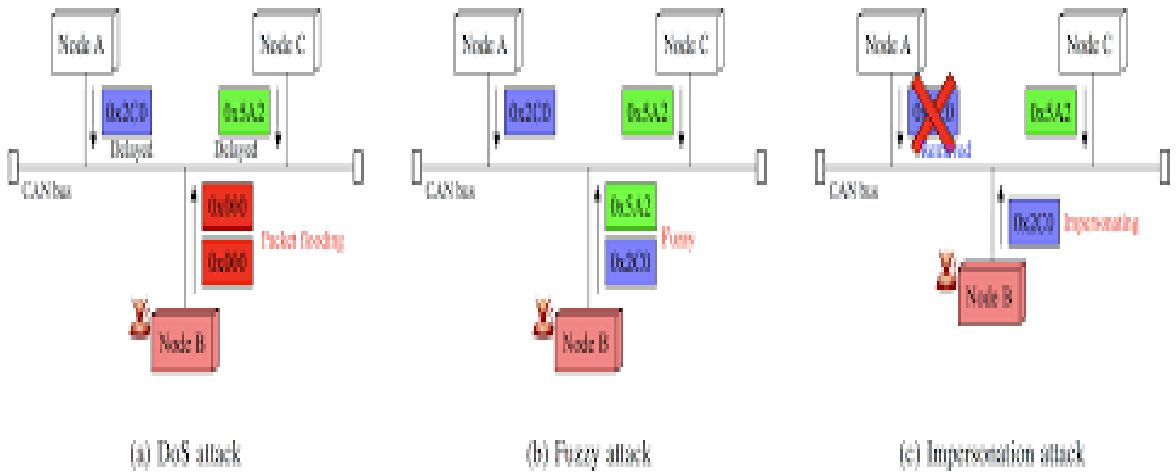

Fig 1: Attacks in dataset

[3]The attacks on dataset could be on resources, network or data. The attacks on resources causes denial of service and hence extra time is consumed while accessing resources. These attacks if severe causes starvation problem. The fuzzy attacks are node based upon the membership functions and if membership functions if abnormal can cause attacks in the dataset. Multiple identity attack is also a problem in which receiver could not determine the sender or vice versa. In all the situations attack hamper the performance of the system considered. This literature focuses on detection of attacks and high degree of classification accuracy.

Rest of the paper is structured as follows: section 2 describes the literature survey on attack detection, section 3 gives the proposed system, section 4 gives the result and performance analysis, section 5 gives the conclusion and future scope and last section gives the references.

II. LITERATURE SURVEY

The literature survey of existing work gives work done towards discovering abnormalities within the dataset that are state driven in nature. The literature survey is presented in terms of comparison table. Table 1 highlight the mechanism used, advantages, disadvantages and parameters used within the existing literature.

Table 1:

Comparative analysis of techniques and parameters of existing literatures

Reference	Technique	Parameters used	Advantages	Disadvantages
[4]	Fuzzy C mean clustering	Prediction accuracy	Prediction accuracy is considerably high	Execution time is a problem that is high
[5]	Virtual data center allocation in cloud	Energy consumption	Energy consumption is minimized	Execution time and false negative rate is high
[6]	Artificial bee colony algorithm for cluster formation	Prediction accuracy and error rate	Using AB colony algorithm, cluster head selection is optimized	Execution time and false negative rate is high
[7]	Fuzzy C mean Clustering	Classification accuracy	Classification accuracy is high using fuzzy c mean clustering approach	False positive rate is high
[8]	Heuristic clustering based approach	Classification accuracy	Accuracy of class selection is high	Error rate is a problem since false negative rate is high
[9]	Fuzzy C mean clustering	Entropy	Degree of relationship between the segment is high	High execution speed is not achieved
[10]	Coverage aware clustering algorithm	Prediction accuracy	Prediction of cluster head selection is high	Error rate could be further minimized by the use of hierarchical clustering procedure
[11]	Classification and clustering algorithm	Error rate	Error rate in cluster head selection for intrusion detection is minimized	Degree of relationship between segments is maximized
[12]	Intrusion detection using feature selection and k means clustering	Prediction accuracy	Prediction accuracy is improved	Error rate can be further minimized using hierarchical clustering procedure

III. PROPOSED SYSTEM

The proposed system uses the hierarchical clustering mechanism to improve the classification accuracy and reduce false positive rate. The false negative rate is minimized by the use of hierarchical clustering procedure. The procedure detects the intrusion with precision and accuracy. The hierarchical clustering procedure builds clusters based on closest pairs. Each clusters is formed by selecting data that is not related with each other. After this closest clusters are merged together. This process continues until no more data is left for distinguishment. The algorithm for the hierarchical clustering is given as under

Algorithm Hierarchical Clustering

- Extract Dataset X={X1,X2,------,Xn}
- Begin with distinct cluster with level 0
- Find the least distance between clusters using Euclidean distance

$$Euc_{dist} = \sqrt{(x - x_i)^2 + (y - y_i)^2}$$

- Update distance matrix according to minimum distance between clusters.
- Stop if all the data points are within the same cluster.
- Predict intruders if data points does not fall within any clusters

The proposed system predict the intruder within prescribed time limits and also gives the least false negative rate with high precision and accuracy.

1. Performance analysis and Results

The performance analysis indicates that the proposed mechanism is better by 10%. The mechanism is implemented using MATLAB 2018b. The result obtained is given as through simulation as

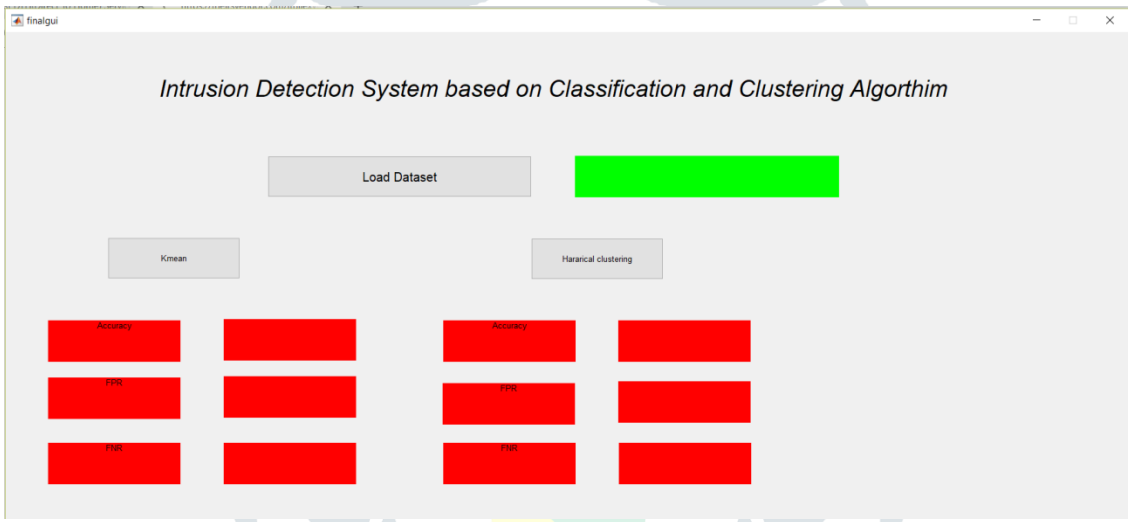

Figure 2: Beginning screen showing the use of intrusion detection mechanism

The mechanism followed is based on k means clustering and hierarchical clustering mechanism. the graphical screen shows the environment where result is obtained in terms of classification accuracy, false positive rate and false negative rate.

The result obtained when user clicks on K-means clustering is given in figure 3

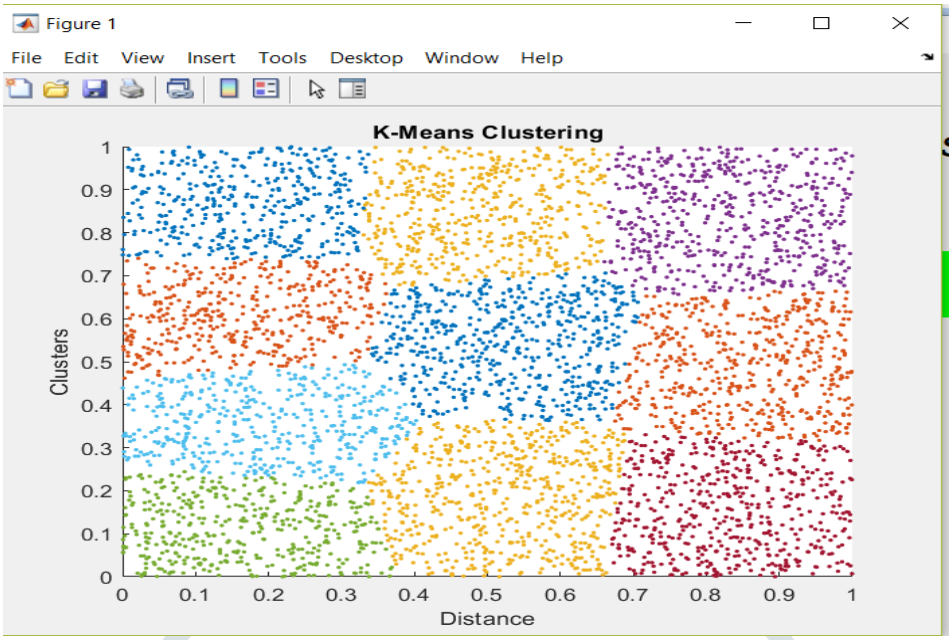

Figure 3:K means clustering procedure with 10 cluster

As k means clustering produce clustering based on value of K which is set to 10 in the proposed literature. Total 10 clusters are produced by the use of k means clustering. The hierarchal clustering result from the simulation is given as under

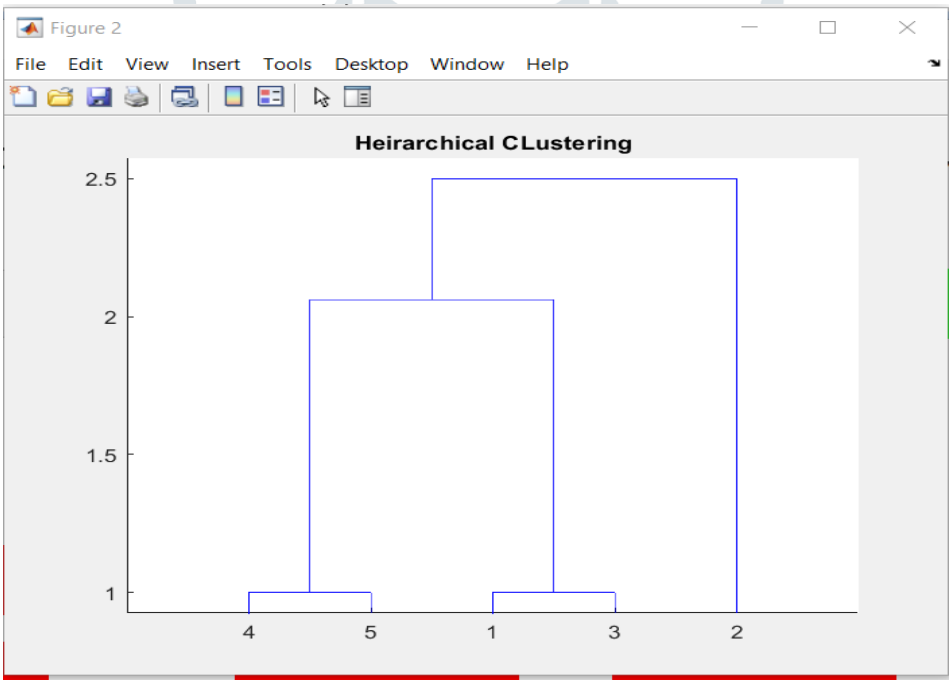

Figure 4: Hierarchical clustering from the dataset

The hierarchical clustering forms only 5 clusters. The clusters so formed are on the basis of relatedness. The mechanism of formation is parent child form where it is possible to express one to many relationship. The obtained result is on the basis of classification accuracy, false positive rate and false negative rate.

The comparative result with k means and hierarchical clustering for the intrusion detection is given as under

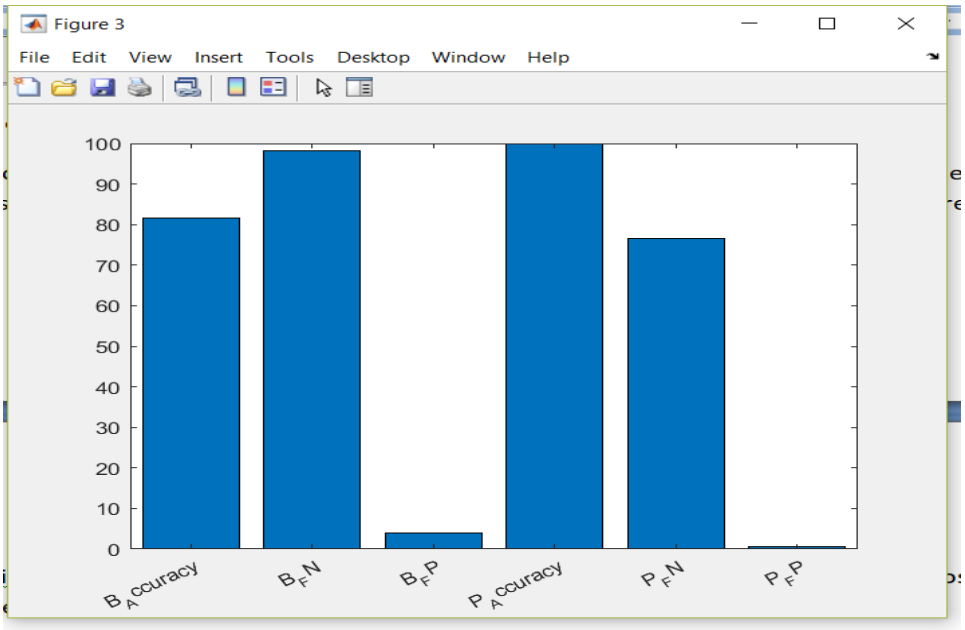

Figure 5: Comparative result for k means and hierarchical clustering

The comparative analysis suggests that hierarchical clustering is better in intrusion detection as compared to k means clustering mechanism.

2. CONCLUSION AND FUTURE SCOPE

The intrusion detection using the application mechanism of clustering is the focus of this literature. The intrusion detection using k means clustering by setting value of k at 10 is accomplished and intrusion is detected with classification accuracy of 81% and with hierarchical clustering the classification accuracy of 91% is achieved. The overall improvement is of 10% which is significant. False detection rate is also decreased. This is measured by the use of false positive rate and false negative rate.

In future, noise handling mechanism can be accommodated with the hierarchical clustering to improve classification accuracy further.

3. REFERENCES

[1] S. Muhammad, S. Hussain, and M. Yousaf, "Neighbor Node Trust Based Intrusion Detection System for WSN," *Procedia - Procedia Comput. Sci.*, vol. 63, pp. 183–188, 2015.

[2] Q. Yan, F. R. Yu, Q. Gong, and J. Li, "Software-defined networking (SDN) and distributed denial of service (DDOS) attacks in cloud computing environments: A survey, some research issues, and challenges," *IEEE Commun. Surv. Tutorials*, vol. 18, no. 1, pp. 602–622, 2016.

[3] N. Alsaedi, F. Hashim, and A. Sali, "Energy Trust System for Detecting Sybil Attack in Clustered Wireless Sensor Networks," *IEEE Access*, no. Micc, pp. 91–95, 2015.

[4] Z. Sun, L. Gao, S. Wei, and S. Zheng, "A Fuzzy C-Means Clustering Algorithm and Application in Meteorological Data," *2010 Second Int. Conf. Model. Simul. Vis. Methods*, pp. 15–18, 2010.

[5] L. Shi, D. Katramatos, and D. Yu, "Virtual Data Center Allocation with Dynamic Clustering in Clouds," *IEEE Access*, 2014.

[6] D. C. TRAN Dang Cong, WU Zhijian, WANG Zelin, "A Novel Hybrid Data Clustering Algorithm Based on Artificial Bee Colony, Algorithm and K-Means," *Chinese J. Electron.*, vol. 24, no. 4, 2015.

[7] S. P. Chatzis, "A fuzzy c-means-type algorithm for clustering of data with mixed numeric and categorical attributes employing a probabilistic dissimilarity functional," *Expert Syst. Appl.*, vol. 38, no. 7, pp. 8684–8689, 2011.

[8] J. Zhao, K. Yang, X. Wei, Y. Ding, L. Hu, and G. Xu, "A Heuristic Clustering-Based Task Deployment Approach for Load Balancing Using Bayes Theorem in Cloud Environment," *IEEE Trans. Parallel Distrib. Syst.*, vol. 27, no. 2, pp. 305–316,

Feb. 2016.

[9] S. Yang, J. Choi, S. Bae, and M. Chung, "A Hybrid Prediction Model Integrating FCM Clustering Algorithm with Supervised Learning," in *IEEE Access*, Springer Singapore, 2015, pp. 619–629.

[10] F. Awad, E. Taqieddin, and A. Seyam, "Energy-Efficient and Coverage-Aware Clustering in Wireless Sensor Networks," *Acm*, vol. 2012, no. July, pp. 142–151, 2012.

[11] S. Kiruthiga, "Detecting Cloning Attack in Social Networks Using Classification and Clustering Techniques," *IEEE*, 2014.

[12] A. Sharma, "Spam Filtering using K mean Clustering with Local Feature Selection Classifier," *ijca*, vol. 108, no. 10, pp. 35–39, 2014.