Efficient Clustering Technique with Feature Reduction Mechanism for Network Anomaly Detection

R. K. Jeyauthmigha
Department of Computer Science &Engineering,
Kongu Engineering College,
Erode, TamilNadu, India.

R. C. Suganthe
Department of Computer Science &Engineering,
Kongu Engineering College,
Erode, TamilNadu, India.

***Abstract*- Today the information technology has grown largely with increased number of users accessing the network, which results in the increasing invasion of security threats. Anomaly detection systems (ADS) monitor the behaviour of a system and flag significant deviations from the normal activity as anomalies. Since any malicious activity on network may lead to serious consequences, the importance of information carried out in these networks makes the task of anomaly detection very crucial. Hence a robust anomaly detection technique, Efficient clustering technique with feature reduction mechanism for network anomaly detection is proposed. The proposed system has two phases. The first phase is used to reduce the dimensions of the dataset. For this, feature reduction method, Recursive feature Elimination technique is employed which selects the relevant features from the dataset while leaving the redundant ones. The Random forest classifier is used to find the Accuracy, Precision, Recall and F-measure of the Recursive feature elimination technique. Then the second phase is the clustering phase which is used to cluster the dataset to classify the labels of the dataset. Here in the proposed system the following clustering techniques are used and the results are compared. The clustering techniques that are evaluated are Hierarchical agglomerative Clustering, Density Based Clustering and K-Means Clustering.**

Keywords- Hierarchical agglomerative Clustering, Density Based Clustering, K-means Clustering, Recursive Feature Elimination Technique.

I.INTRODUCTION

Anomaly detection is the problem of finding any deviations in data that do not conform to normal behavior. These nonconforming patterns are often referred to as anomalies, outliers, discordant observations, exceptions, aberrations, surprises, peculiarities, or contaminants in different application domains, where anomalies and outliers are used in the context of anomaly detection[1].

Anomaly detection finds extensive use in a wide variety of applications such as fraud detection for credit cards, insurance, or health care, intrusion detection for cyber-security, fault detection in safety critical systems, and military surveillance for enemy activities [1].

The most widely deployed methods for detecting cyber terrorist attacks and protecting against cyber terrorism employ signature-based detection techniques. Such methods can only detect previously known attacks that have a corresponding signature, since the signature database has to be manually revised for each new type of attack that is discovered. These limitations have led to an increasing interest in intrusion detection techniques based on data mining [2, 3, 4, 5, 6].

Clustering techniques have been used successfully to the anomaly detection problem, where it is applied to the normal samples to generate a set of clusters that will represent the normal class. Clustering is an unsupervised learning technique of data mining that takes unlabeled data points and tries to group them according to their similarity: points assigned to the same cluster have high similarity, while the similarity between points assigned to different clusters is low [7]. When a clustering algorithm deals with noisy information, the algorithm is called robust [8], [9], [10], and when the number of clusters is determined automatically, it is usually called unsupervised [11].

The proposed work uses data mining techniques to find anomaly data objects. Data mining techniques are classified as three types. They are Supervised, Semi-Supervised and Unsupervised techniques. Supervised technique uses the class labels to identify a data object. Semi-Supervised technique is capable of detecting anomalies either by class labels or without class labels. The Unsupervised technique can identify a data object without the class labels.

II. LITERATURE SURVEY

Xie M et al. [12] analyzed a few of the key design principles relating to the development of anomaly detection techniques in WSNs .Then, the state-of-the-art techniques of anomaly detection in WSNs are systematically introduced, according to WSNs' architectures (Hierarchical/Flat) and detection technique categories (statistical techniques, rule based, data mining, computational intelligence, game theory, graph based, and hybrid, etc.).

Mohiud din Ahmed et al. [13] surveyed a detailed study of four primary group of anomaly detection techniques which include statistical, classification, clustering and information theory. The paper brings down that clustering and information theory based techniques are better than other techniques.

Martin Ester et al. [14] analyzed the performance of DBSCAN using synthetic data and real data of the SEQUOIA 2000 benchmark. Experiments demonstrated that DBSCAN is significantly more effective in discovering clusters of arbitrary shape than the well-known algorithm CLARANS.

Markus M. Breunig et al. [15] gave a detailed formal analysis showing that Local Outlier Factor has many desirable properties. Experiments performed using real world datasets demonstrated that Local Outlier Factor can be used to find outliers.

Mohammad Reza Parsaei et al. [16] studied that the datasets used in intrusion detection are not balanced and the ability of detecting two attack classes, R2L and U2R, is lower than that of the normal and other attack classes. In order to overcome this issue, employed a hybrid approach. This hybrid approach is a combination of synthetic minority oversampling technique (SMOTE) and cluster center and nearest neighbour (CANN). Relevant features are selected by leave one out method (LOO). Moreover, this study employs NSL KDD dataset. Results indicate that the proposed method improves the accuracy of detecting U2R and R2L attacks in comparison to existing systems by 94% and 50%, respectively.

Dr. Y. P. Raiwani and Shailesh Singh Panwar [17] have analyzed four different clustering algorithms using NSL-KDD dataset and tried to cluster into two classes i.e. normal and anomaly using K-Means, EM, DB clustering and COBWEB. The aim of the valuation is to decide the class labels of different type of data present in intrusion detection dataset and to find out efficient clustering algorithm. Results show that K-Means outperforms in time and accuracy to classify the dataset.

Riti Lath Manish and Shrivastava [18] analyzed data using different techniques i.e. K-Means which is based on clustering, and k-nearest neighbour, support vector machine are classification techniques and concluded that after evaluation k-nearest neighbour gives better result than SVM for classifying normal and abnormal data but it takes more time for its execution. The work done proposed that classification technique is better in a particular case than clustering, as K-Means fails to give better separation.

Michael Steinbach et al. [19] resented the results of an experimental study of some common document clustering techniques: agglomerative hierarchical clustering and K-Means. The results shows that the bisecting K-Means technique achieves better results than the standard K-Means approach and better than the hierarchical approaches that were tested.

Lance Parsons et al. [20] performed a survey of the various subspace clustering algorithms along with a hierarchy organizing the algorithms by their defining characteristics and compare the two major approaches to subspace clustering using empirical scalability and accuracy tests and discussed some potential applications where subspace clustering could be particularly useful.

Shahreza ML et al. [21] proposed a novel method for detecting traffic anomalies in a network by extracting and capturing their features in the transform domain. It considers topology and traffic of the network. It finds the high frequency nature of the traffic of the network. This motivates to utilize transform domain analysis theory to characterize network-wide traffic to identify its abnormal components. Besides, grouping all origin–destination flows in the network in accordance with common destination nodes is carried out. By using the network information of the topology and transform-domain analysis in the given time window, the spurious traffic components can be found and identified. Simulation results show that detection algorithm is better than the previous algorithms.

Yang X.S and Deb S [22] formulate a new meta-heuristic algorithm, called Cuckoo Search (CS), for solving optimization problems. This algorithm is based on the brooding behavior of some cuckoo species in combination with the Levy flight behavior of some birds.

III. PROPOSED METHODOLOGY

A. Overview

In this chapter, the approach of the proposed work is discussed in detail. The constructed anomaly detection system works in two phases namely feature reduction phase and clustering phase which is discussed in detail below. In the feature reduction phase the number of features are reduced to extract the relevant features from the actual dataset. Then in the clustering phase the dataset are trained for classifying the labels. The proposed system is devised to find the best clustering technique to find anomalies of the three clustering techniques namely K-means clustering, Hierarchical agglomerative clustering and Density based clustering.

B. Workflow of the proposed methodology

The *Fig. 1* explains the working flow of the devised work. Firstly the dataset is cleaned without the redundant and irrelevant data. Then the dataset is applied for the Recursive feature elimination method for extracting the relevant features. Then the reduced feature subset dataset is given as input to the clustering techniques for classifying the labels. The results are compared and the optimized clustering technique is finally found to detect the anomalies.

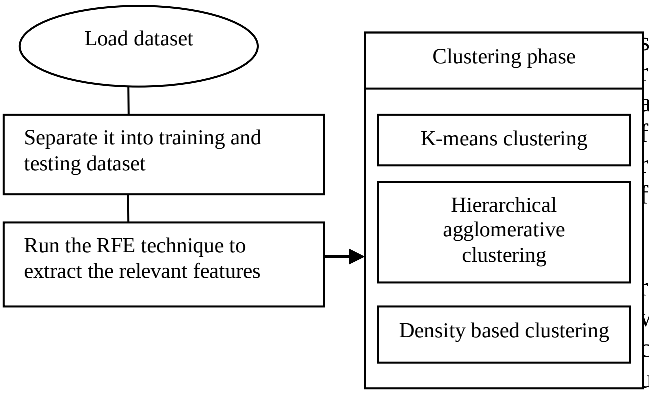
Load dataset
Separate it into training and testing dataset
Run the RFE technique to extract the relevant features
Clustering phase
K-means clustering
Hierarchical agglomerative clustering
Density based clustering

Fig. 1. Block diagram of the proposed methodology

C. *KDD Dataset description*

The KDD Cup 1999 dataset used for benchmarking intrusion detection problems is used in the experiment. The dataset was a collection of simulated raw TCP dump data. KDD training dataset consists of approximately 4,900,000 single connection vectors each of which contains 41 features and is labeled as either normal or an attack, with exactly one specific attack type. The simulated attacks fall in one of the following four categories.

1) Denial of Service Attack (DoS): It is an attack in which the attacker makes some computing or memory resource too busy or too full to handle legitimate requests, or denies legitimate users access to a machine[23].

2) User to Root Attack (U2R): It is a class of exploit in which the attacker starts out with access to a normal user account on the system (perhaps gained by sniffing passwords, a dictionary attack, or social engineering) and is able to exploit some vulnerability to gain root access to the system[23].

3) Remote to Local Attack (R2L): It occurs when an attacker who has the capability to send packets to a machine on a network but who does not have an account on that machine exploits some vulnerability to gain local access as a user of that machine[23].

4) Probing Attack: It is an attempt to gather information about a network of computers for the apparent purpose of circumventing its security controls[23].

1) Feature Reduction Phase

In the feature reduction phase, the Recursive feature elimination method which is a wrapper method is used. The uses of feature reduction phase are as follows:

i. It facilitates the algorithm to train faster.

ii. It diminishes the intricacy of the model and makes it easier to understand.

iii. It improves the accuracy of the model if the consistent subset of features is selected.

i) Recursive Feature Elimination Technique

A huge datasets with large dimensions always induce severe impacts to anomaly detection techniques. The reduced dataset with relevant features achieve high accuracy than the actual dataset. Hence the recursive feature elimination method is proposed to remove the redundant and irrelevant features and extract the relevant features from the actual dataset.

It aims to find the best performing feature subset. It repeatedly creates models and keeps aside the best or the worst performing feature at each iteration. It then constructs the next model with the remaining features until all the features are exhausted. It then ranks the features based on the order of their elimination. The test of hypothesis performed in this technique is the analysis of variance (ANOVA) that is appropriate to compare the means of a continuous variable in two or more independent comparison groups. The fundamental strategy of ANOVA is to consistently check variability within groups being compared and also check variability among the groups being compared.

The accuracy, precision, recall and F-measure of this recursive feature elimination technique are evaluated using the random forest classifier.

2) Clustering Phase

In the clustering phase, the three clustering techniques namely K-means clustering, Hierarchical agglomerative clustering and Density based clustering are used. They are then compared by the results obtained, to find the efficient to find anomaly data objects.

ii) K-Means Clustering Algorithm

K-Means Clustering Algorithm is a partitioning algorithm. It find solutions to the clustering problems. It form clusters with the given data objects of the dataset in such a way that the distance between the data objects inside the cluster is smaller than the data objects outside the cluster. The sum of squared distance measure is calculated to each of the data objects inside the cluster. The Euclidean distance is computed for each data object and the distance to which cluster is minimum is chosen to allocate the data object to its cluster. Thus clusters are formed. The Euclidean distance is computed by the equation (1) given below,

$$Distance(d,c) = \sqrt{\sum_{i=0}^{N} \sum_{j=1}^{k} (p_i - c_j)^2} \qquad (1)$$

where p_i is the point at which the data object lies and c_j is the cluster to which the data object belongs to with the total number of N data objects.

The K-Means algorithm is designed to work in two phases. The first phase is the initialization phase in which the initial centroids are initialized and the second phase is the iterative phase where the final centroids for the cluster is achieved by iterative calculation of the centroid value of the cluster by using the Euclidean distance until there is no change in the centroid points.

Algorithm:

Step:1 Cluster the data into k groups where k, no. of clusters is predefined.
Step:2 Select k points at random as cluster centers.
Step:3 Assign objects to their closest cluster center according to the Euclidean distance function.
Step:4 Update the cluster centroid.
Step:5 Repeat steps 2 ,3 and 4 until the same points are assigned to each cluster in consecutive rounds.

ii) Hierarchical Agglomerative Clustering

Hierarchical clustering algorithms are either top-down or bottom-up. Bottom-up approach is called agglomerative clustering. It successively merges pairs of clusters taking the data objects in the dataset one by one on the basis of the nearest distance measure of all the pair wise distance between the data object. Then the distance between the data object is recalculated and the distance that is to be calculated when the clusters has been formed are of five types. They are single linkage, complete linkage, average linkage, centroid distance, ward's method.

In single-link (or single linkage) hierarchical clustering, smallest minimum pair wise distance between the two data objects in the clusters is considered. In complete-link (or complete linkage) hierarchical clustering, the two clusters are merged with the maximum pair wise distance from one data object from one cluster to another data object of another cluster. In average link hierarchical clustering, the two clusters are merged with the mean distance between elements of each cluster. Centroid linkage uses the Euclidean distance between the centroids of the two clusters.

In ward's method hierarchical clustering, the two clusters are merged with the minimum in variance for the cluster being merged. An hierarchical agglomerative clustering can be visualized as a dendrogram. A dendrogram is a diagram used to demonstrate the arrangement of the clusters produced by hierarchical clustering.

PSEUDOCODE:

```
Let X = {x1, x2, x3, ..., xn} be the set of data points.
Form n clusters each with one element
Construct a graph T by assigning one vertex to each cluster
while there is more than one cluster
•   Find the closest clusters C1 and C2
•   Merge C1 and C2 into new cluster C with |C1|+|C2| elements
•   Compute distance from C to all other clusters
if they are close
•   Add a new vertex C to T and connect to vertices C1 and C2
•   Remove rows and columns of d corresponding to C and C2
•   Add a row and column to d corresponding t the new cluster C
  return T
```

iii) Density Based Clustering

The density based clustering method (DBSCAN) is based on clustering the dataset by the density properties found in the data objects. DBSCAN stands for Density-Based Spatial Clustering of Applications with Noise. The clustered dataset consists of an increased density of data objects clustered together which belongs to a single class and the data objects that are dispersed are called outliers as they do not belong to any cluster as they have low densities.

The density based algorithm requires at most two parameters: a density metric, MinPts and the minimum size of a cluster, Eps ε. In density based clustering, the data objects are classified as core points, border points and outliers. A point is a core point if it has more than a specified number of points (MinPts) within Eps ε and these are points that are at the interior of a cluster. A point is a border point which has minimum number of points than the MinPts within Eps ε, but is in the closest distance of a core point. A noise point (outlier) is any point that is not a core point or a border point and these are points not reachable from any other point.

DBSCAN is efficient even when applied on large databases and calculating the number of clusters apriori is not necessary. If the number of data objects that belong to a cluster are less than the minimum number of points, MinPts then those data objects do not form clusters. They remain as outliers. Thus the cluster is formed by the given two parameters Eps ε and MinPts. By Fine-tuning those values, one can achieve clusters of varying shapes and densities.

PSEUDOCODE:

```
DBSCAN(D, epsilon, min_points):
 C = 0
 for each unvisited point P in dataset
     mark P as visited
     sphere_points = regionQuery(P, epsilon)
     if sizeof(sphere_points) < min_points
         ignore P
     else
         C = next cluster
         expandCluster(P, sphere_points, C, epsilon, min_points)
       add P to cluster C
for each point P’ in sphere_points
     if P’ is not visited
         mark P’ as visited
         sphere_points’ = regionQuery(P’, epsilon)
```

```
            if sizeof(sphere_points’) >= min_points
                sphere_points = sphere_points joined with
sphere_points’
            if P’ is not yet member of any cluster
                add P’ to cluster C
    regionQuery(P, epsilon):
    return all points within the n-dimensional sphere
```

IV. RESULTS AND DISCUSSIONS

A. Result Analysis

The KDD dataset is employed in the proposed system and the feature reduction method, recursive feature elimination technique is applied to the dataset. Initially the dataset consists of 41 features and after applying this technique the number of features reduced to 13 which are the relevant features to classify the labels in the dataset. The accuracy, precision, recall, F-measure are evaluated by using the random forest classifier and the results are generated. The results obtained is 98% accuracy, 97.3% precision, 97.1% recall, 97% F-measure.

The Clustering techniques employed in the proposed work are compared for their proficiency in Homogeneity, Completeness, V-measure, Adjusted Rand Index and Adjusted Mutual Information. The results obtained showed that the Density based Clustering (DBSCAN) showed supremacy of the other two clustering techniques.

B. Confusion Matrix

A table layout that is used to visualize the performance of an algorithm. The Table.I shows the confusion matrix table. Each row of the matrix represents the instances in a predicted class while each column represents the instances in an actual class. It is a table with two rows and two columns that reports the number of false positives, false negatives, true positives, and true negatives.

TABLE. I. CONFUSION MATRIX

	Predicted positives	**Predicted negatives**
Actual positives	**True positives(TP)**	**False positives(FP)**
Actual negatives	**True negatives(TN)**	False negatives(FN)

- **True positives:** The activity is intrusive and is reported as anomalous
- **True negatives:** The activity is not intrusive and is not reported as intrusive
- **False positives:** The activity is not intrusive but the intrusion detection system reports it as intrusive. These are called false positives because an intrusion detection system falsely reports intrusions
- **False negatives:** These are intrusive but reported as not anomalous. An intrusion detection system fails to detect this type of activity as anomaly. These are called false negatives because the intrusion detection system falsely reports the absence of intrusions

C. Evaluation metrics

The various evaluation metrics used are discussed below.

1. **Detection Rate**: Detection Rate (Recall) is the measure of the completeness of the classifier. It is the ratio between the true positive to the true positive and false negative.

$$\text{DetectionRate}=\frac{TP}{(TP+FN)} \qquad (2)$$

2. **Accuracy**: Accuracy is used as a statistical measure of how well a binary classification test correctly identifies or excludes a condition. That is, the accuracy is the proportion of true results (both true positives and true negatives) among the total number of cases examined. It is the measure of the classifier to predict the correctness of the algorithm model.

$$\text{Accuracy}=\frac{(TP+TN)}{(TP+TN+FP+FN)} \qquad (3)$$

3. **Precision**: It is the measure of the exactness of the classifier. It is the number of correct results divided by the number of all returned results. It is the measure of the true positive to the true positive and false positive. Precision is also referred to as positive predictive value

$$\text{Precision}=\frac{TP}{(TP+FP)} \qquad (4)$$

4. **False Positive Rate**: The false positive rate is calculated as the ratio between the number of negative instance wrongly categorized as positive (false positives) and the total number of actual negative instances (regardless of classification).It is the misclassified measure of the true positives.

$$\text{False Positive Rate}=\frac{FP}{(FP+TN)} \qquad (5)$$

5. **F-Measure**: F-measure is the harmonic mean of Recall and Precision It is to enhance the model against either precision or Detection Rate.

$$\text{F-Measure}= 2 * \frac{(Detection\ Rate*Precision)}{(Detection\ Rate+ Precision)} \qquad (6)$$

6. **Homogeneity:** It is a measure of the ratio of samples of a single class pertaining to a single cluster. A clustering result satisfies homogeneity if all of its clusters contain only data points which are members of a single class.

$$H(C)=-\sum_{c=1}^{|C|}\frac{\sum_{k=1}^{|K|}a_{ck}}{N}\log\frac{\sum_{k=1}^{|K|}a_{ck}}{N} \qquad (7)$$

7. **Completeness:** It measures the ratio of the member of a given class that is assigned to the same cluster. A clustering result satisfies completeness if all the data points that are members of a given class are elements of the same cluster.

$$H(K)=-\sum_{k=1}^{|K|}\frac{\sum_{c=1}^{|C|}a_{ck}}{N}\log\frac{\sum_{c=1}^{|C|}a_{ck}}{N} \qquad (8)$$

8. **V-measure:** It is the harmonic mean of homogeneity and completeness, expressed by the following formula,

V = 2 * (homogeneity * completeness) / (homogeneity + completeness) (9)

9. **Adjusted Rand Index**: The Rand Index computes a similarity measure between two clusters by considering all pairs of samples and counting pairs that are Rand index adjusted for chance assigned in the same or different clusters in the predicted and true clusters.

10. **Adjusted Mutual Information**: It is an adjustment of the Mutual Information (MI) score to account for chance. It accounts for the fact that the MI is generally higher for two clusters with a larger number of clusters, regardless of whether there is actually more information shared.

TABLE. II. CLASSIFICATION RESULTS OF CLUSTERING TECHNIQUES

Clustering Techniques	True positive (TP)	True negative (TN)	False positive (FP)	False negative (FN)
K.MEANS	2182	452	1983	0
HAC	1983	2145	489	0
DBSCAN	2634	1979	4	0

The Table. II shows the True positive (TP), True negative (TN), False positive (FP), False negative (FN) results obtained by the clustering techniques namely K-means, Hierarchical agglomerative clustering, Density based clustering.

The Table. III tabulates the results for the clustering techniques namely K-means clustering, Hierarchical agglomerative clustering (HAC), Density based clustering (DBSCAN). The clustering techniques are evaluated for the quality of clusters using the metrics Homogeneity, Completeness, V-measure, Adjusted Rand Index and Adjusted Mutual Information. Then it is also measured for the accuracy, precision, recall, F-measure and false positive rate. Thus the Density based clustering is found to be the efficient clustering technique for handling the high dimensional datasets.

TABLE. III. PERFORMANCE EVALUATION OF CLUSTERING TECHNIQUES

Clustering Techniques	Homogeneity	Completeness	V-Measure	Adjusted Rand Index	Adjusted Mutual Information	Accuracy	Recall	Precision	F-Measure	False Positive Rate
K.MEANS	100	72.3	83.9	81.5	72.3	90.2	90.2	82.8	90.6	1.85
HAC	100	71.4	83.3	80.4	71.4	89.4	80.2	99.9	88.9	1.85
DBSCAN	100	99.1	99.5	99.9	99.1	99.9	99.8	99.9	99.8	0.2

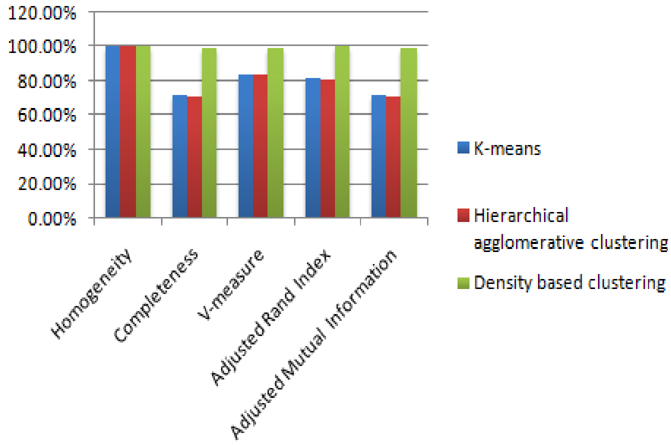

Fig. 2. Comparison of Clustering Techniques

The *Fig. 2* shows the evaluated results of the clustering techniques for the metrics Homogeneity, Completeness, V-measure, Adjusted Rand Index and Adjusted Mutual Information. The evaluated results for the K-means clustering are 100%, 72.3%, 83.9%, 81.5%, and 72.3% respectively. For the Hierarchical agglomerative clustering the results are 100% Homogeneity, 71.4% Completeness, 83.3% V-measure, 80.4% Adjusted Rand Index and 71.4% Adjusted Mutual Information. For the Density based clustering the results estimated are 100% Homogeneity, 99.1% Completeness, 99.5% V-measure, 99.9% Adjusted Rand Index and 99.1% Adjusted Mutual Information.

A good clustering technique should have a low false positive rate and a high detection rate (recall).The clustering techniques which are discussed are thus measured for accuracy, precision, detection rate (recall), F-measure and false positive rate. The results evaluated showed that the Density based clustering is having the lowest false positive rate and the highest detection rate (recall).

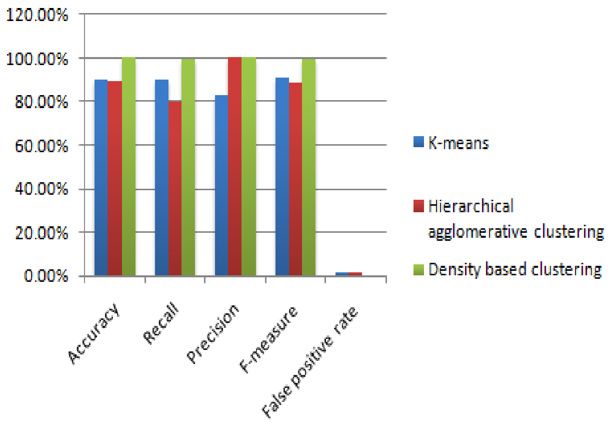

Fig. 3. Comparison of the Performance of Clustering Techniques

The *Fig. 3* shows the results of the clustering techniques in terms of accuracy, precision, recall, F-measure and false positive rate. For the K-means clustering the estimated results are 90.2% accuracy, 90.2% precision, 82.8% recall, 90.6% F-measure, 1.85% false positive rate. For the Hierarchical agglomerative clustering the calculated results are found to be 89.4% accuracy, 80.2% precision, 99.9% recall, 88.9 F-measure, 1.85% false positive rate. For the Density based clustering the evaluated results are 99.9% accuracy, 99.8% precision, 99.9% recall, 99.8% F-measure, 0.2% false positive rate.

V. CONCLUSION

The anomaly detection technique, implemented aims to identify attacks or malicious activity in a network with a high detection rate while maintaining a low false positive rate. It was executed in two phases namely feature reduction phase and clustering phase. In the former phase, recursive feature elimination method is implemented to reduce the number of features. The reduction is based on selecting the relevant features while leaving the irrelevant and redundant features in the actual dataset. Here, random forest classifier is used to evaluate the feature reduction method. Then the clustering techniques are applied in the reduced feature subset dataset. Here the goal is to achieve the best clustering technique of the three techniques namely K-means clustering, Hierarchical agglomerative clustering and Density based clustering implemented. For that the evaluation metrics namely Homogeneity, Completeness, V-measure, Adjusted Rand Index and Adjusted Mutual Information, accuracy, precision, recall, F-measure and false positive rate are measured and compared. The results estimated showed that the Density based clustering is capable of handling high dimensional dataset with high Detection Rate and low False Positive Rate with high quality of clustering the dataset.

The future work is to use the ensemble classifiers to improve the accuracy for the clustering techniques that is to be implemented while dealing with high dimensional network dataset with multi objective functions.

REFERENCES

[1] Varun Chandola , Aridam Banerjee , Vipin Kumar “Anomaly Detection: A Survey”, ACM Computing Surveys, Vol. 41, No. 3, Article 15, 2009.

[2] W. Lee, S. J. Stolfo, Data Mining Approaches for Intrusion Detection, Proceedings of the 1998 USENIX Security Symposium, 1998.

[3] E. Bloedorn, et al., Data Mining for Network Intrusion Detection: How to Get Started, MITRE Technical Report, 2001.

[4] J. Luo, Integrating Fuzzy Logic With Data Mining Methods for Intrusion Detection, Master's thesis, Department of Computer Science, Mississippi State University, 1999.

[5] D. Barbara, N. Wu, S. Jajodia, Detecting Novel Network Intrusions Using Bayes Estimators, First SIAM Conference on Data Mining, Chicago, IL, 2001.

[6] S. Manganaris, M. Christensen, D. Serkle, and K. Hermix, A Data Mining Analysis of RTID Alarms, Proceedings of the 2nd International Workshop on Recent Advances in Intrusion Detection (RAID 99), West Lafayette, IN, 1999.

[7] R. Duda and P. Hart, “Pattern Clarrfication and Scene Analysis”, NY:Wilev Interscience, 1973.

[8] R.Krishnapuram and J.M Keller, “A Possibilistic Approach to Clustering”, IEEE Trans. Fuzzy Syst. Vol.13,pp.791-802, 1997.

[9] J. M. Jolion, P. Meer. and S. Bataouche, “Robust clustering with applications in computer vision”, vol. 13, pp. 791-802, 1991.

[10] 0. Nasraoui and R. Krishnapum, “Clustering using a genetic fuzzy least median of squares algorithm” in North American Fuzzy Informtion Processing Society Conference, (Syracuse NY), 1997.

[11] 0. Nasraoui and R. Krishnapuram, “A novel approach to unsupervised robust clustering using genetic niching” in proceedings of the Ninth IEEE International Conference on Fuzzy Systems, pp. 170-175, 2000.

[12] M. Xie , S. Han , B. Tian and S. Parvin , “Anomaly detection in wireless sensor networks: a survey”, J NetwComput Appl, 2011.

[13] M. Ahmed , A.N. Mahmood and J. Hu, “A survey of network anomaly detection techniques”, J NetwComput Appl, 2016.

[14] Martin Ester, Hans Peter Kriegel, Jorg Sander, Xiaowei Xu, A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”, kdd Proceedings, 1996.

[15] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, Jorg Sander, “LOF: Identifying Density-Based Local Outliers”, Conf. On Management of Data, 2000.

[16] Mohammad Reza Parsaei , Samaneh Miri Rostami ,Reza Javidan, “A Hybrid Data Mining Approach for Intrusion Detection on Imbalanced NSL-KDD Dataset”, International Journal of Advanced Computer Science and Applications, 2016.

[17] Dr. Y. P. Raiwani, Shailesh Singh Panwar, “Research Challenges and Performance of Clustering Techniques to Analyze NSL-KDD Dataset”, International Journal of Emerging Trends & Technology in Computer Science, 2014.

[18] Riti Lath, Manish Shrivastava, “Analytical Study of Different Classification Technique for KDD Cup Data’99”, International Journal of Applied Information Systems, 2012.

[19] Michael Steinbach, George Karypis, Vipin Kumar, “A Comparison of Document Clustering Techniques”, Technical Report 00-034, 2000.

[20] Lance Parsons, Ehtesham Haque, Huan Liu, “Subspace Clustering for High Dimensional Data: A Review”, ACM SIGKDD Explorations Newsletter, 2004.

[21] M.L. Shahreza , D. Moazzami , B. Moshiri and M. Delavar ,”Anomaly detection using a self-organizing map and particle swarm optimization”, ScientiaIranica, 2011.

[22] Yang X-S, Deb S, “Cuckoo search via Levy flights”, In: Proceedings of World Congress on Nature & Biologically Inspired Computing. IEEE Publications, 2009.

[23] Mahbod Tavallaee , Ebrahim Bagheri , Wei Lu , and Ali A. Ghorban , “A Detailed Analysis of the KDD CUP 99 Data Set”, IEEE Symposium on Computational Intelligence in Security and Defense Applications, 2009.