International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 9 Issue 7, July-2020, Impact Factor: 6.754

Intrusion Detection System by using Fuzzy Algorithms

Shahbaa I. Khaleel

Software Dept., College of Computer Science and Mathematics, Mosul University, Iraq

ABSTRACT

Intrusion Detection System (IDS) is an important component in the framework of computer and information security and its main goal is to distinguish between the system's regular activities and the behaviors that we can describe as intrusion. Last defense in the system. The project introduces some methods of fuzzy techniques to detect network anomalies: The first method is the Kmeans clustering algorithm conventional algorithm. Second method The hard weighted kmeans clustering algorithm was built. The third method is to use the fuzzy kmeans clustering algorithm. These system were implemented on the US Air Force local network simulation data, the KDD Data set, which consists of a set of data, which is the training data of 494020. The performance of the intrusion detection system was measured in this research using several measures to detect intrusion and it was found that the best results were obtained from the method of the fuzzy clustering algorithm.

Keywords: intrusion detection, clustering algorithm, fuzzy logic, kmeans , hard kmeans, fuzzy kmeans.

1. Introduction

The complexity of distributed computer systems, their importance, and the information resources available in them have grown very quickly, based on this fact. Computers and their networks have become targets of computer crime, which have increased more and more [1].The Internet user will participate in many applications that need secure transfers within the computer network when transferring information and financial credits, as banks and commercial markets that rely on the Internet every day work to achieve access to protected password sites. As these movements increase, the system will become vulnerable to attacks that store processes and important data [2].IDS provides two main benefits: Visibility and Control. In the case of combining these benefits, it is possible to create and strengthen the security policy of the organization working on the security of the private network. Visibility is the ability to observe and understand the nature of packet traffic within a computer network, while Control is the control of packet traffic in a computer network and includes access to the computer network or a part of it. Clarity represents the supreme authority in decision-making and makes it possible to configure the security policy adopted in measuring it against real data. As for control, it is the key that puts implementation and makes it possible to interact with security policy [3].

In this research, fuzzy clustering techniques were used in the intrusion detection process. Initially, the traditional cluster algorithm of the kmeans algorithm was used, then the hard kmeans algorithm was used, which is the starting point for the fuzzy cluster. and then using fuzzy kmeans clustering algorithm that it's best of all algorithms.

2. Intrusion Detection Systems IDS

Intrusion Detection Systems IDS are very important for anyone who wants to protect their information or devices from theft or to keep the information confidential and secret. It detects the presence of intruders if any breach is made to alert the user to the necessary precautions either by locking programs or operating the protection device and so on, without these programs And the systems may not pay attention to the user who penetrates the device and perform operations that help the hacker and make it easier for him to detect and steal confidential information. At the present time, the technologies used in the internet are sophisticated and manifold and some of them may use them to penetrate and attack and may cause massive damage to devices, networks and funds[4].To know the goals of the saboteurs, there are many and varied goals. The goal may be theft, vandalism, or obstruction, or it may be only for intrusion and revealing secret matters. This has harmed many people and companies and the cause of many financial losses. Protection and guarding systems for important information and devices, and also for networks used to connect to the Internet. One of these devices and systems is the intrusion detection system[5].In previous years, the intrusion detection system, IDS, became one of the hottest researches in computer security. It is an important detection technology that is used as a countermeasures to maintain integrity and data availability during intrusion. Intrusion techniques may include exploiting software errors, system errors, password breaches, or exploiting design errors in specific protocols, so the intrusion detection system is a system that exposes these breaches and reports them accurately to the appropriate authorities. Intrusion detection systems are user-specific and are an important tool in implementing the information security system for companies and organizations, which includes defining rules and practices for providing security, handling breaches, and recovering from damage caused by security breaches [6].

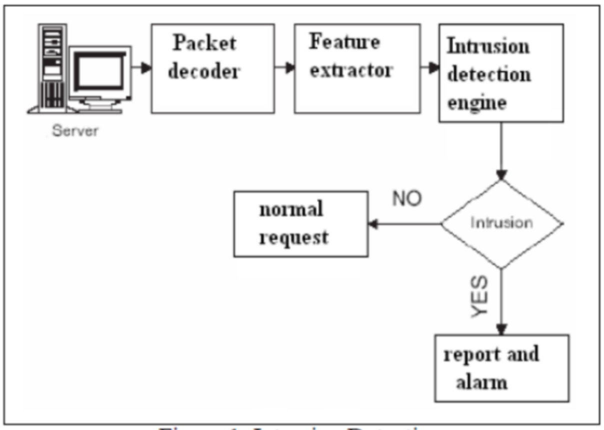


Figure.1. Intrusion Detection

3. Clustering

Ensemble clustering is a promising approach that combines the results ofmultiple clustering algorithms to obtain a consensus partition by merging different partitionsbased upon well-defined rules[7].Clustering methods partition a set of objects into clusters such that objects in the same cluster are more similar to each other than objects in different clusters according to some defined criteria[8].Clustering is an unsupervised learning problem, which tries to group a set of points into clusters such that points in the same cluster are more similar to each other than points in different clusters, under a particular similarity metric [9].The goal of clustering is to identifying the clusters, which can be considered as classes[10]. Clustering can be used to produce an effective image index as follows: After clustering, each cluster is represented by its centroid or sometimes a single representative data item and, instead of the original data items, the query point is compared to the centroids or the cluster representatives. The best cluster or clusters, according to the used similarity measure, are then selected and the data items belonging to those clusters are retrieved also according to the used similarity measure [11,12].

3.1 K-means Clustering Algorithm

The k-mean algorithm is the most frequently used clustering algorithm due to its simplicity and efficiency. K-means is a partitional clustering algorithm. It performs iterative relocation to partition a dataset into k cluster [13]; and it is based on the minimization of a performance index which is defined as the sum of the squared distances from all points in a cluster domain to the cluster center. This algorithm consists of the following steps: [14,15]

Step 1: Choosing K initial cluster centers $z_1(1), z_2(1), ..., z_K(1)$. These are arbitrary and are usually selected as the first K samples of the given sample set.

Step 2: Distributing the samples $\{x\}$ at the k^{th} iterative step among the K cluster domains, using the relation:

$$x \in S_j(k) \quad if \quad \|x - z_j(k)\| < \|x - z_i(k)\|$$

for all $i = 1,2,...,K, i \neq j,$ where $S_j(k)$ denotes the set of samples whose cluster is $z_j(k)$.

Setp 3: Computing the new cluster centers $z_j(k+1), j = 1,2,...,K$, such that the sum of the squared distances from all points in $S_j(k)$ to the new cluster center is minimized. In other words, the new cluster center $z_j(k+1)$ is computed so that the performance index:

$$J_j = \sum_{x \in S_j(k)} \|x - z_j(k+1)\|^2 \qquad , j = 1,2,...,K$$

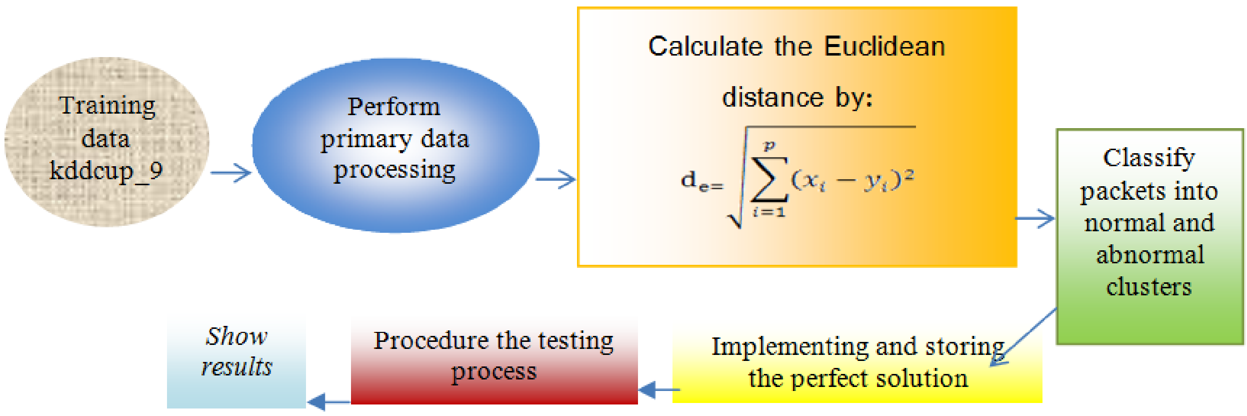
is minimized. The $z_j(k+1)$ which minimizes this performance index is simply the sample mean of $S_j(k)$. Therefore, the new cluster center is given by: $z_j(k+1) = \frac{1}{N_j} \sum_{x \in S_j(k)} x \quad , i = 1,2,\ldots,K$

where, N_j is the number of samples in $S_j(k)$. The name "K-means" is obviously derived from the manner in which clusters are sequentially updated.

Step 4: If $z_j(k+1) = z_j(k)$ for $j = 1,2,\ldots,K$, the algorithm has converged and the procedure is terminated. Otherwise one should go to step2.

The behavior of the k-means algorithm is influenced by the number of cluster centers specified, the choice of initial cluster centers, the order in which the samples are taken, and, of course, the geometrical properties of the data. Although no general proof of convergence exists for this algorithm, it can be expected to yield acceptable results when the data exhibit characteristic pockets which are relatively far from each other. In most practical cases the application of this algorithm will require experimenting with various values of K as well as different choices of starting configurations [16].

In this research, the kmeans clustering algorithm was used, where the cluster centers are chosen with the number of required classifications, which are normal and abnormal, then the traditional distance for each packet of data is calculated with the centers of the first and second clusters where if the first distance is less or equal to the second distance means that the package belongs to the cluster The first, and if the first distance is greater than the second distance, it means that the packet belongs to the second cluster and so on for all the packets.The following figure show the steps of this algorithm.



4. Fuzzy Clustering

Fuzzy clustering is a process of assigning the membership levels, and then using them to assign data elements to one or more clusters. It gives more information on the similarity of each object[17].Fuzzy algorithms can assign data to multiple clusters. Fuzzy algorithms are based on fuzzy logic [18] or fuzzy clustering is an extension of the concept of fuzzy logic which was developed by LotfiZadeh. In fuzzy logic, Zadeh suggested that a statement could be partially true and partially false [19]. Deterministic membership functions assign each datum to a particular cluster, that is, the membership functions map the membership of the datum in a cluster to either 0 or 1. Fuzzy membership functions, on the other hand, map the memberships to the real interval [0.0 1.0]. The degree of membership in the cluster depends on the closeness of the datum to the cluster center. High membership values indicate less distance between the datum and the cluster center[20,21].Clustering involves the process of arranging data points in such a way that items sharing similar characteristics are grouped together.

4.1 Hard weighted k-means Algorithm

The primary goal of this Hard weighted kmeans algorithm is to define each point in the data space with a cluster by relying on the segmentation of the C-Partition of the data space X into a set of clusters $\{C_i, i = 1,2,\ldots, c\}$.If we assume that the data set $X = \{x_1, x_2, \ldots, x_n\}$ is a finite set of data points (vectors), and that c is the number of clusters, then we will notice that it is Exhaustive, meaning that its union is equal to the data space [22,23] X. That is: $\bigcup_{i=1}^{c} C_i = X$ Likewise, it is evident that these clusters are mutually exclusive (Disjoint), meaning that there is no intersection between them. That is, the following two restrictions are achieved:

i) $C_i \cap C_j = \phi$ for all i≠j

ii) $\phi \subset C_i \subset X$, all i

The last limitation states that the cluster cannot be empty and is definitely a subset of the data space X, and as it is noticed, the number of clusters c is confined between 2 and n, i.e :For example, if c = n is one of the clusters, then each data point for its own cluster will be placed alone, while if c = 1, then all data points will be placed in the same cluster. The HKM algorithm needs to pre-determine the number of clusters, that is, enter the number of clusters required from the start.The HKM algorithm aims to find the center in each cluster and reduce the Objective Function (also called the Dissimilarity Function [2]) to measure the distance, and the traditional distance is chosen as a function of the distance, because this function depends on the distance $d(\mathbf{x}_k - \mathbf{c}_i)$ between the data point x_k and the center of the cluster c_i, The objective function can be expressed by the following equation:

$$J = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{k,\mathbf{x}_k \in C_i} d(\mathbf{x}_k - \mathbf{c}_i) = \sum_{i=1}^{c} \left(\omega\beta \sum_{k,\mathbf{x}_k \in C_i} \left\| \mathbf{x}_k - \mathbf{c}_i \right\|^2 \right)$$

Whereas, J_i is the objective function within cluster i, and $\omega$is a *k*-by-*m* matrix, and $\omega^{\beta}$is the weight of the *l*th dimension in the *i*th cluster. $\beta$is a parameter that greater than 1.and symbolizes the Euclidean distance which is calculated as in the following equation:

$$\text{Euclidean :dis}(\mathbf{t}_i, \mathbf{t}_j) = \sqrt{\sum_{h=1}^{k} (t_{ih} - t_{jh})^2}$$

The segmented clusters can be defined as a c*n two-dimensional U matrix which is called the Membership Matrix, where each entry in the U array is referred to as u_{ik}. If the x_k data point belongs to cluster i then u_{ik} = 1, otherwise it is zero. So u_{ik} can be expressed by the following equation:

$$u_{ik} = \begin{cases} 1 \ \mathit{if} \ \left\| \mathbf{x}_k - \mathbf{c}_i \right\|^2 \le \left\| \mathbf{x}_k - \mathbf{c}_j \right\|^2 \ , \ \mathit{for\ each}\ \ j \ne i \\ \\ 0 \ \mathit{otherwise} \end{cases}$$

And the U matrix has the following two properties:

1- The sum of the elements of each column equals one. That is:

$$\sum_{i=1}^{c} u_{ik} = 1 \text{ ; for all k=1, 2, …, n}$$

2- The sum of all elements of the matrix U equals n. That is:

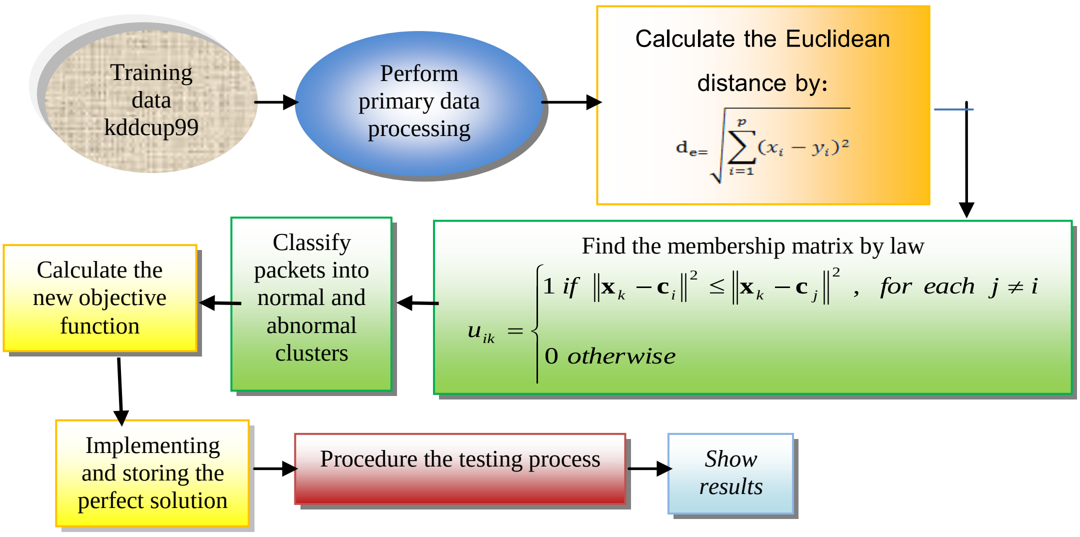
$$\sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik} = n$$

The optimal center, ci, which represents the average center, Centriod, is the average of all data points in cluster i, according to the following formula:

$$\mathbf{c}_i = \frac{1}{|C_i|} \sum_{k,\mathbf{x}_K \in C_i} \mathbf{x}_k$$

The x_k data point belongs to the cluster C_i if the center ci is the center closest to that point among all the centers. K is the number of data points in cluster C_i.The input to this algorithm is to specify the number of clusters required, enter data in vectors, and enter the Threshold value. The outputs are the clusters in which the elements are grouped, the membership matrix, the center matrix, the number of iterations, and the objective function.

In this research, a hard weighted k-meansclustering algorithm was used, where the cluster centers are chosen by the number of classifications that we want, which are normal and abnormal. Then, the traditional distance for each packet of data is calculated with the centers of the first and second clusters where if the first distance is less or equal to the second distance means that the packet belongs For the first cluster, the largest affiliation is given to the membership function, which is one. If the first distance is greater than the second distance, it means that the packet belongs to the second cluster and so on for all packets.The lowest affiliation is given to the membership function, which is zero, and the distance between the centers of the clusters and the packet is multiplied by the weight of the fuzzy value.The following figure show the steps of hard algorithm.



4.2 FuzzyK-meansAlgorithm

The cluster algorithm that was mentioned earlier produces fragile clusters, which means that the data point either belongs to the cluster or does not belong to it, and thus the clusters are not overlapping. This type of fragmentation is called fragile cluster[24].The task of the cluster leads to the presentation of algorithms that use the concepts of fuzzy logic, and these algorithms will classify the results as fuzzy clusters, and this means that the data point will belong to several clusters at one time and with different Membership Grade (degree) memberships. And the most important algorithm belonging to this class is the Fuzzy K-Means (FKM) algorithm which is an extension of the old K-Means algorithm but in fuzzy logic applications[25]. The FKM algorithm was proposed by Researcher Bezdek in 1981. This algorithm is a cluster technology separated from HKM, as the HKM algorithm is mainly based on the philosophy of the Crisp Set, which uses a hard segmentation of data points, so that the affiliation of these points is definitively determined: either that it belongs to a particular cluster or not, or an algorithm FKM is based on the philosophy of fuzzy logic that mainly depends on the idea of gradual and not sharp affiliation, so the FKM algorithm differs from HKM, that is, the FKM algorithm uses Fuzzy Partitioning that allows each data point to belong to a cluster with a specific degree of membership degree and thus all a data point can belong to several about us A registration at the same time and with different membership levels falling in the period (1, 0),[26,27]. If we assume that the data set $X = \{x_1, x_2, ..., x_n\}$ is a finite subset of the set of real numbers. Suppose that c is the number of clusters, which is an integer such that $2 \le c \le n$, so the FKM algorithm splits the data set X into c from the fuzzy clusters, so that the data in the same group is as similar as possible and differs in different groups as possible. Thus, the fuzzy partitioning of the X dataset can be represented by the membership matrix U of dimensional c*n, since each entry in the U matrix is referred to as u_{ik} and is within the Range: $u_{ik} \in [0,1], \forall i = 1,...,c, \forall k = 1,...,n$

The U Membership Matrix has two main properties as described above. The goal of FKM is similar to what exists in HKM, which is finding the center for each cluster and reducing the objective function, which can be represented by the following equation[28]:

$$J(\mathbf{U}, \mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik}{}^m)(d_{ik})^2$$

Since J_i is the target function within cluster i. D_{ik} is the traditional distance between the data point x_k and the center ci and is calculated by an equation.

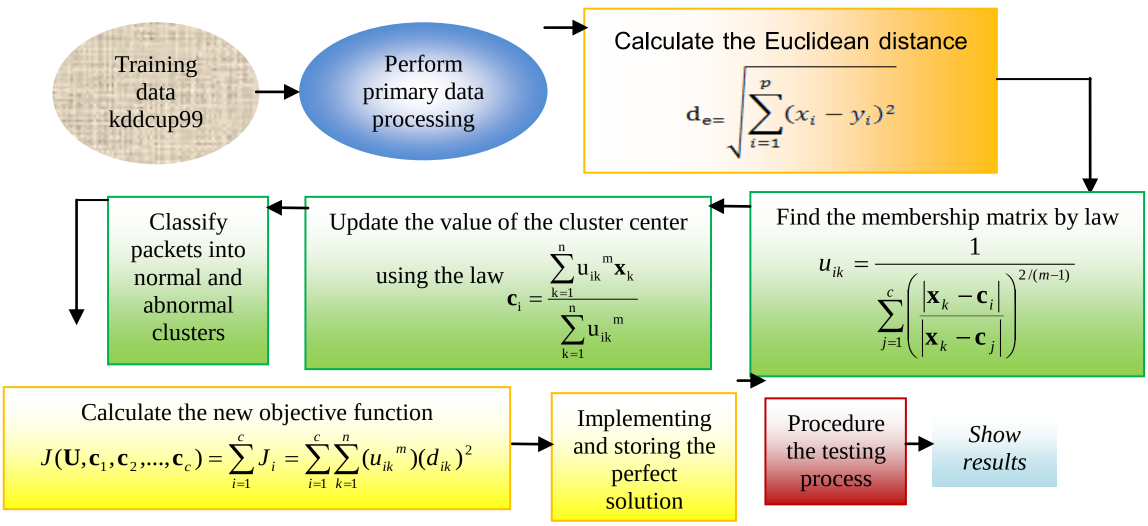
$$\mathbf{c}_i = \frac{\sum_{k=1}^{n} u_{ik}{}^m \mathbf{x}_k}{\sum_{k=1}^{n} u_{ik}{}^m} \quad \text{for all}$$

Whereas, c_i is the center of the cluster i, and i = 1,2, ..., c.

$$u_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{|\mathbf{x}_k - \mathbf{c}_i|}{|\mathbf{x}_k - \mathbf{c}_j|}\right)^{2/(m-1)}} \quad \text{for all k}$$

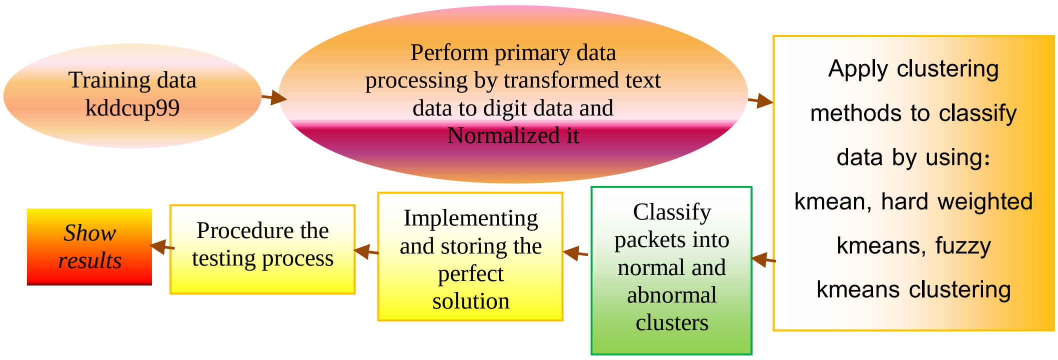
The value of m is the weighted power of the membership and is a real number and it is used to control the fuzzing of the membership for each Datum maintenance, and there is no basis for optimal selection of m. But m = 2 is usually chosen. In our field of research, we also depended on choosing the value of m = 2. The value of m deserves additional interests. When $m \rightarrow 1$, the solution to the FKM procedure advances to Hardness or Crispness (meaning that the degree of membership of the data point becomes close to 1 or 0), while the largest value of m (for example 2 or more) will result in fogging Membership (for example, the degrees of membership for all clusters are close to each other).Cannon (1986) and others have suggested that when choosing any value of m is between 1.3 and 1.8 we will get high execution performance toward cutting. Suggested Nikhil and Bezdek, (1995) states that when choosing the value of m in the period [1.5, 2.5], it gives a correct cluster result [29,30].The input to this algorithm is to enter the data in vectors, choose the number of clusters c, determine the threshold value, and determine the value of m. The outputs of this algorithm are clusters with a number c that includes the elements after their classification, the number of iterations, the membership matrix U, the center matrix, and the objective function.

In this research fuzzy kmeans clustering algorithm is used to classify the packets, here cluster centers are selected with the number of required classifications, which are normal and abnormal, then the membership function is calculated to obtain the centers of new clusters, then the Euclidean distance is calculated for each packet of data with the centers of the first and second clusters and give the degree of affiliation of each of the packets entered for the membership function where the number of elements is calculated Depending on the first and second cluster, by the max function, and then refreshing the value of the membership function.All previous operations are repeated until the stopping condition is reached to reach the optimal solution or until the specified number of cycles iteration stop is completed and then the optimal solution (cluster centers) is stored. The following figure show the steps of this algorithm.



5. Experimental and Results of Proposed System

A proposed intrusion detection system was built and designed using traditional and fuzzy clustering methods to conduct detection. The stages of the system begin with the initial processing process of data, as the entries for this process are the KDD99 text file for the training process, which consists of 494020 lines of text in addition to the KDD99 text file for the test process, which consists of 311029 lines of text and contains new attacks that are not present in the test file to measure Efficiency of the system after the training process. It has a Normalization process where data is normalized and its values are converged, where the values are all between (0-1) and then the outputs of this process are KDD99 digital files with consistent data ready for training and testing.



Initially the traditional clustering algorithm was used where in this process the algorithm is trained to detect parasitism by training data and to find the optimal solution and training is done in a number of iterations. And then the testing process is conducted where in this process the efficiency of the optimal solution obtained at the training stage is tested. And by using the optimal solution that represents the centers of the clusters in separating the test data using the law of Euclidean Distance and distributing the data on the clusters, this process is repeated with the number of test data and then the intrusion detection measures are calculated to ensure the efficiency of the optimal solution.

Table 1 show the results of kmeans method

Measures & information	Training process	Testing process
Iterationsstop	30	1
Iterationnum	10	1
Packetnum	100000	100000
Normal	59064	59998
Abnormal	40936	40002
Targetnormal	56237	60593
Targetabnormal	43763	39407
FP	0	595
FN	2827	0
TP	40936	40002
TN	59064	59998
FNR	0.06459795	0
FPR	0	0.00981962
TNR	1	0.99018
TPR	0.9354021	1
Precision	1	0.985344
Accuracy	0.9725072	0.994085
detectionrate	97.173000	99.405
Execution time (M:S)	0:20.389331	0:4.726830

In the proposed system, the hard weight kmeans method was used to detect of intrusion, and the following results were obtained:

Table 2 show the results of hard weight kmeans method

Measures & information	Training process	Testing process
Iterationsstop	30	1
Iterationnum	10	1
Packetnum	100000	100000
Normal	59064	59998
Abnormal	40936	40002
Targetnormal	56237	60593
Targetabnormal	43763	39407
FP	0	595
FN	2827	0
TP	40936	40002
TN	59064	59998
FNR	0.06459795	0.99018
FPR	0	0
TNR	1	1
TPR	0.9354021	0.00987962
Precision	1	0.985344
Accuracy	0.9725072	0.994085
detectionrate	97.173000	99.405
Execution time (M:S)	0:18.064916	0: 0.702005

Also, the fuzzy clustering method that represents of fuzzy kmeans was used to detect intrusion in the packets entering the system and this method was better than the previous two methods. The following results were obtained.

Table 3 show the results of fuzzy kmeans method

Measures & information	Training process	Testing process
Iterationsstop	10	1
Iterationnum	7	1
Packetnum	100000	100000
Normal	56418	59635
Abnormal	43582	40365
Targetnormal	56237	60593
Targetabnormal	43763	39407
FP	0	958
FN	181	0
TP	43582	40365
TN	56418	59635
FNR	0.004135914	0.98419
FPR	0	0
TNR	1	1
TPR	0.9958641	0.0158104
Precision	1	0.990511
Accuracy	0.9981933	0.976817
detectionrate	99.819000	99.042000
Execution time (M:S)	0:6.1	0:0.3

6. CONCLUSIONS

In this research we designed and implemented the intrusion detection system,After applying the intrusion detection methods used in the research on the KDDcup99 data set, the following was observed: With regard to the first method of the traditional clustering algorithm that represented by kmeans algorithm, it gave good results in intrusion detection, and this was shown by calculating the values of system performance measures such as the detection ratio and the percentage of positive and negative errors and other special measures By detecting intrusion. As for the second method that represented by hard weighted kmeans, it gave better results than the first method, depending on the membership function and the weight of the degree of membership.As for the third method represented by the fuzzy clustering algorithm, both the objective function and the membership function were relied upon and the best results were obtained when they were trained on the same training data for the previous two methods. This method achieved the highest detection rate among the previous methods.

REFERENCES

[1] KusumBharti,et al.,2010,Fuzzy K-mean Clustering Via J48 For Intrusiion Detection System, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 1 (4) , 2010, 315-318.

[2] Al-Taey, S. I. 2006. Efficient content-based multimedia retrieval with neural network, Ph.D. thesis, Computer science-Faculty of Computers & Mathematics Sciences, University of Mosul, Iraq.

[3] Shahbaa I. Khaleel, Dr. Nidhal Hussein, Image Retrieval Based on Content by using Clustering Algorithms , International Journal of Enhanced Research in Science, Technology & Engineering, ISSN: 2319-7463, Vol. 8 Issue 7, July-2019. DOI:http://www.doi.org/ste/19.0703

[4] Manar Y. Kashmola, Bayda I. Khaleel, 2012, Clustering and Detecting Network Intrusion Based on Fuzzy Algorithms, AL-Rafidain Journal of Computer Sciences and Mathematics Volume 9 , Issue (2), Pages 125-138 , Publisher: Mosul University, DOI: 10.33899/csmj.2012.163706.

[5] Shahbaa I. Khaleel, Karammohammedmahdisaleh, 2012, Detection of network anomaly based on hybrid intelligence techniques, AL-Rafidain Journal of Computer Sciences and Mathematics Volume 9 , Issue (2), Pages 81-98 , Publisher: Mosul University, DOI: 10.33899/ csmj.2012.163720.

[6] Manar Y. Ahmed, Bayda I. Khaleel, 2013, Using Artificial Intelligence Techniques For Intrusion Detection System, AL-Rafidain Journal of Computer Sciences and Mathematics Volume 10 , Issue (1), Pages 157- 175, Publisher: Mosul University, DOI: 10.33899/csmj.2013.163433. Fifth Scientific Conference Information Technology 2012 Dec. 19-20.

[7] HarunPirim , et al. , 2011,PERFORMANCE OF AN ENSEMBLE CLUSTERING ALGORITHM ON BIOLOGICAL DATA SETS, *Mathematical and Computational Applications*, Vol. 16, No. 1, pp. 87-96, 2011.

[8] Ali Salem Bin Samma and Rosalina Abdul Salam, 2009,Adaptation of K-Means Algorithm for Image Segmentation, World Academy of Science, Engineering and Technology 50 2009.

[9] Dr. Shahbaa Ibrahim Khaleel, Color Image Compression Based on Fuzzy Clustering , International Journal of Enhanced Research in Science, Technology & Engineering, ISSN: 2319-7463, Vol. 8 Issue 7, July-2019. DOI:http://www.doi.org/ste/19.0702.

[10] Basu S., (2003), Semi-Supervised Clustering: Learning With Limited User Feedback, PhD Thesis, University of Texas, Austin.

[11] Krishnamachari S., Abdel-Mottaleb M., (1999), Hierarchical Clustering Algorithm For Fast Image Retrieval, Conference on Storage and Retrieval For Image and Video Databases, pp. 427-435.

[12] BaydaaIbraheemKhaleel, 2014,Image Clustering based on Artificial Intelligence Techniques, AL-Rafidain Journal of Computer Sciences and Mathematics Volume 11 , Issue (1), Pages 99-112 , Publisher: Mosul University, DOI: 10.33899/csmj.2014.163735.

[13] Dr. Shahbaa I. Khaleel, Dr. Nidhal Hussein, Content Based Video Retrieval Using Back propagation Neural Network and Radial Basis Function Network , International Journal of Enhanced Research in Science, Technology & Engineering, ISSN: 2319-7463, Vol. 8 Issue 12, December -2019. DOI:http://www.doi.org/ste/19.1202.

[14] Shahbaa I. Khaleel, Baydaa I. Khaleel, Alaa I. khaleel, Image Compression Based on Artificial Intelligent Techniques, AL-Rafidain Journal of Computer Sciences and Mathematics Volume 6 , Issue (3), Pages 75-109 , Publisher: Mosul University, DOI: 10.33899/csmj.2009.163839.

[15] Nada N. Seleem, Baydaa I. Khaleel, Shahbaa I. Khaleel, Artificial Intelligent Techniques with Watermarking, AL-Rafidain Journal of Computer Sciences and Mathematics Volume 6 , Issue (2), Pages 229-266 , Publisher: Mosul University, DOI: 10.33899/csmj.2009.163810.

[16] Tou J., Gonzalez R., (1974), Pattern Recognition Principles, Addision-Wesley Publishing Company, USA.

[17] P. Valarmathie, et al.,2009, Hybrid Fuzzy C-Means Clustering Technique for Gene Expression Data, International Journal of Research and Reviews in Applied SciencesISSN: 2076-734X, EISSN: 2076-7366Volume 1, Issue 1(October 2009).

[18] Dr. ManarYounis K., Dr. BaydaIbraheemKhaleel, 2013, Hybrid Fuzzy and Neural Network for Intrusion Detection System, *JOURNAL OF EDUCATION AND SCIENCE*, Volume 26, Issue 4, Pages 152-170, Publisher: Mosul University, DOI: 10.33899/edusj.2013.89971.

[19] AL-Taey, baydaaIbraheemKhaleel, 2012, Artificial Intelligent Techniques for Intrusion Detection and Classification, Ph.D. Thesis, Computer Science, College of Computer Sciences and Mathematics, University of Mosul.

[20] Dr. Shahbaa I. Khaleel, Dr. Nidhal Hussein, Colored Image Retrieval Based on Content by Using Integration of Features for Many Color Space Models , International Journal of Enhanced Research in Science, Technology & Engineering, ISSN: 2319-7463, Vol. 8 Issue 8, August-2019. DOI:http://www.doi.org/ste/19.0803

[21] Dr. Shahbaa I. Khaleel, Dr. Nidhal Hussein, Color Image Retrieval Based on Content by Using Linear and Nonlinear Combiner, International Journal of Enhanced Research in Science, Technology & Engineering, ISSN: 2319-7463, Vol. 8 Issue 8, August-2019. DOI:http://www.doi.org/ste/19.0802.

[22] Shailendra S. , PremNarayan A. , (2012), Comparison of K-means and Modified K-mean algorithms for Large Data-set , International Journal of Computing, Communications and Networking, Volume 1, No.3, pp. 106-110 .

[23] Shahbaa I. Khaleel and Ragadwaleedkhaled, 2014, Selection and Prioritization of Test Cases by using Bees Colony, AL-Rafidain Journal of Computer Sciences and Mathematics Volume 11 , Issue (1), Pages 179-201 , Publisher: Mosul University, DOI: 10.33899/csmj.2014.163746.

[24] Audrey P Gasch and Michael B Eisen,2002,Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering, 2002 Gasch and Eisen, licensee BioMed Central Ltd(Print ISSN 1465-6906; Online ISSN 1465-6914).

[25] Shahbaa I. Khaleel, Ashraf Abdulmunim Al Thanoon, 2013, Design a Tool for Generating Test Cases using Swarm Intelligence, AL-Rafidain Journal of Computer Sciences and Mathematics Volume 10 , Issue (1), Pages 421-444 , Publisher: Mosul University, DOI: 10.33899/csmj.2013.163468.

[26] Taoying Li and Yan Chen,2010, Fuzzy *K*-Means Incremental Clustering Based on *K*-Center and Vector Quantization, JOURNAL OF COMPUTERS, VOL. 5, NO. 11, NOVEMBER 2010,pp. 1670-1677.

[27] Baydaa I. Khaleel, 2014, Using Artificial Intelligence Techniques for Image Compression, AL-Rafidain Journal of Computer Sciences and Mathematics Volume 11 , Issue (2), Pages 65-81 , Publisher: Mosul University, DOI: 10.33899/csmj.2014.163750.

[28] Binu Thomas and Raju G,2009,A Novel Fuzzy Clustering Method for Outlier Detection in Data Mining, International Journal of Recent Trends in Engineering, Vol. 1, No. 2, May 2009, pp. 161-165.

[29] Fardos Adnan Abdalkader, Shahbaa I. Khaleel, Nada N. Saleem, 2010, Hybrid intelligent watermark System, AL-Rafidain Journal of Computer Sciences and Mathematics Volume 7 , Issue (2), Pages 125-138 , Publisher: Mosul University, DOI: 10.33899/csmj.2013.163902.

[30] Disha Sharma ,2011,Fuzzy Clustering as an Intrusion Detection Technique, International Journal of Computer Science & Communication Networks,Vol 1(1),September-October 2011, pp. 69-75.