Anomaly based Intrusion Detection using Modified Fuzzy Clustering

B. S. Harish and S. V. Aruna Kumar *

Department of Information Science and Engineering, Sri Jayachamarajendra College of Engineering, Mysuru, Karnataka (India)

Received 22 March 2015 | Accepted 3 May 2017 | Published 12 May 2017

Abstract

This paper presents a network anomaly detection method based on fuzzy clustering. Computer security has become an increasingly vital field in computer science in response to the proliferation of private sensitive information. As a result, Intrusion Detection System has become an indispensable component of computer security. The proposed method consists of three steps: Pre-Processing, Feature Selection and Clustering. In pre-processing step, the duplicate samples are eliminated from the sample set. Next, principal component analysis is adopted to select the most discriminative features. In clustering step, the network samples are clustered using Robust Spatial Kernel Fuzzy C-Means (RSKFCM) algorithm. RSKFCM is a variant of traditional Fuzzy C-Means which considers the neighbourhood membership information and uses kernel distance metric. To evaluate the proposed method, we conducted experiments on standard dataset and compared the results with state-of-the-art methods. We used cluster validity indices, accuracy and false positive rate as performance metrics. Experimental results inferred that, the proposed method achieves better results compared to other methods.

Keywords

Anomaly Detection, Intrusion Detection, Fuzzy Clustering, Principal Component Analysis, Robust Spatial Kernel Fuzzy C-Means.

DOI: 10.9781/ijimai.2017.05.002

I. Introduction

Computer security has become an increasingly vital field in computer science in response to the proliferation of private sensitive information. The term "Intrusion" refers to any unauthorized access which attempts to compromise confidentiality, integrity and availability of information resources [1] [14] [32]. Traditional intrusion prevention techniques such as firewalls, access control and encryption have failed to fully protect systems from sophisticated attacks. As a result, Intrusion Detection System has become an indispensable component of computer security which is used to detect the aforementioned threats. In 1987, Denning [7] first proposed an intrusion detection model. Since then many researchers have been focusing on developing efficient and accurate Intrusion Detection System (IDS) models. The intrusion detection techniques fall under two types, Misuse or Signature Based and Anomaly Based methods. Signature based methods detect only known intrusion attacks whose signatures are stored in the database. These methods fail to detect unknown intrusions. On the other hand, anomaly based methods detect the attacks based on the signature deviation.

In early days, intrusion detection is done using rule based approaches, where experts define a set of rules for normal and abnormal conditions. These systems work better for known attacks but fail to detect unknown attacks. In later 1990's researchers concentrated to develop automatic intrusion detection methods. Many researchers used data mining and machine learning algorithms to detect unknown attacks. Among various intrusion detection techniques, Fuzzy Logic based methods play a very important role. From literature review it is found that clustering methods are widely used approaches in intrusion detection system. Jianliang et al, [13] developed an intrusion detection system using K-means clustering algorithm. The experimentation was carried out on standard KDD-99 dataset. Cluster to class mapping, No class and Class Dominance are the key problems in K-means clustering. To overcome these drawbacks, Bharti et al., [4] developed two variants of traditional K-means algorithm. Ren et al., [23] developed a Fuzzy C-Means (FCM) algorithm to detect intrusions. The intrusion detection model was built through carrying out fuzzy partition and clustering of data. The experimental result shows that the algorithm can effectively separates normal and abnormal data. To overcome cluster centre initialization and convergence problem of FCM, Wang et al., [29] proposed a hybrid algorithm for intrusion detection system. This hybrid method combines FCM with Quantam behaved Particle Swarm Optimization. The Particle Swarm Optimization algorithm is used to overcome the drawback of FCM and to achieve global optimization and fast convergence. Guorui et al., [10] developed a semi supervised Fuzzy C-Means clustering algorithm for intrusion detection. This method overcomes the drawbacks of FCM i.e Sensitivity to the initial values and converging to the local minima by using few labelled data to improve the learning ability of the Fuzzy C-Means. Sampat and Sonawani [25] developed an intrusion detection system using Improved Dynamic Fuzzy C-Means (IDFCM) clustering. The IDFCM is a variant of the traditional FCM which adaptively updates the cluster centres. Experimental result shows the IDFCM gives better detection accuracy rate than traditional FCM. Hameed et al., [11] developed an hybrid clustering algorithm for intrusion detection. This hybrid algorithm combines Modified Fuzzy Possiblistic C-Means (MFPCM) and symbolic fuzzy clustering. This method uses 30 features with optimal sensitivity and highest discriminatory power.

Ganapathy et al., [9] proposed a intrusion detection system based on Weighted FCM and Immune Genetic Algorithm (GA). The Weighted FCM is a modification of FCM which builds a system for more accurate

* Corresponding author.

E-mail addresses: bsharish@sjce.ac.in (B. S. Harish), arunkumarsv55@gmail.com (S. V. Aruna Kumar).

attack detection. Immune GA improves the performance, probability of gaining the global optimal values and solves the high dimensionality problem. Khazaee and Rad [17] developed a novel method based on Fuzzy C-Means for improving the intrusion detection performance. The experiments were conducted on KDD Cup-99 dataset. Kumar et al., [19] developed a novel method of intrusion detection which involves fuzzy feature clustering. In this method fuzzy features clustering method is used to reduce the dimensionality of system calls. Pandeeswari and Kumar [21] developed hybrid detection system for cloud environment. This hybrid model works in three steps. In the first step, to improve the learning capability of ANN, fuzzy clustering is used. In the second step, various ANN modules are trained according to their cluster values. In the last step, fuzzy aggregation module is used to combine the results of various ANN.

Karthik and Nagappan [16] developed an intrusion detection system using Kernel Fuzzy C-Means and Bayesian Neural Network. This hybrid model consists of two step. In the first step, Fuzzy Bisector Kernel Fuzzy C-means is used to obtain the cluster centers. In the second step the centroids obtained from previous step are used for the learning of Bayesian network. The experiments were conducted on standard KDD Cup 99 Dataset. Rustam and Talita [24] proposed an intrusion detection algorithm based on Fuzzy Kernel C-Means (KFCM). Khazaee and Faez [18] developed an hybrid classification method for network intrusion detection. This hybrid model combines fuzzy clustering with multilayer perceptron neural network. Here training samples are initially clustered using fuzzy clustering and the inappropriate data will be detected and moved to another dataset. Further, Multilayer Perceptron will be trained using new labels. To classify outlier fuzzy ARTMAP neural network is employed. Surana., [27] developed an hybrid algorithm which combines Fuzzy C-Means and Neural Network for intrusion detection. This approach divides the training data into smaller groups using Fuzzy C-Means. Later, Neural Networks are trained using these subsets. Finally individual neural network results are aggregated. Kumar and Harish [2] proposed Robust Spatial Fuzzy C-Means algorithm which considers spatial information and uses kernel distance metric. Xie et al., [31] developed an intrusion detection using hybrid clustering method. This hybrid method is the combination of Fuzzy C-Means Clustering, Average Information Entropy, Support Vector Machine and Fuzzy Genetic algorithm.

It is evident from the above discussion that, the researchers have been attempting to come out with more efficient and robust intrusion detection techniques. Further, most of the existing methods are based on supervised and unsupervised learning approaches. Supervised learning requires a large volume of training samples and they fail to detect unknown attacks. On the other hand, unsupervised learning has got the advantage of detecting unknown attacks. On the contrary, the main limitations of the unsupervised learning are as follows: higher False Alarm Rate (FAR), fails to identify the specific type of attacks and worstly affected by curse of dimensionality. Further, most of the existing unsupervised methods uses euclidean as a distance metric. Unfortunately, euclidean metric is very sensitive to noise which results in degrading the system accuracy.

With this backdrop, to eliminate the above said problems in this paper, we present network anomaly detection method based on fuzzy clustering. The proposed method consists of three steps: Pre-Processing, Feature Selection and Clustering. To evaluate the proposed method, we conducted experiments on standard intrusion detection dataset and compared the results with other variants of FCM methods.

In summary, the main contributions of this paper are as follows:

- A Modified Fuzzy Clustering method (RSKFCM) for anomaly detection is presented. The method can also identify a specific type of attacks.
- To handle curse of dimensionality, we employed Principal Component Analysis (PCA) as feature selection method.
- To overcome the existing drawbacks of FCM methods, we considered neighborhood information which in turn reduces the false alarm rate and increases the system accuracy.
- We used Gaussian kernel as distance measure to compute the distance between cluster center and samples. The advantage of using Gaussian kernel is that it reduces the effect of noise.
- We validated the proposed method on standard intrusion dataset using four cluster validity functions, accuracy, and false alarm rate. Further, we also compared our results with the contemporary methods.

The rest of the paper is organized as follows: Section 2 presents details about the KDD Cup 99 Dataset. Traditional Fuzzy C-Means is presented in Section 3. Section 4 presents the proposed method. Experimental setup, dataset used for experimentation and results are presented in section 5. Conclusions are drawn in section 6.

II. KDD Cup 99 Dataset

Since 1999, Many researchers evaluated the intrusion detection models on the KDD Cup 99 dataset [26]. This dataset was originally created by 1998 DARPA intrusion detection evolution program. The dataset contains five and two millions of training and testing samples respectively. Each sample as a set of 41 features derived from each connection and a label which specifies connection as normal or attack. These 41 features can be categorized into 4 groups.

- Basic features: These features can be derived from packets header without inspecting the payload.
- Content features: These features contains the information present in the payload of the original TCP packets and extracted using domain knowledge.
- Time based Traffic features: These features contains the service information which inspect the connection in the past two seconds that have the same service as the current connections.
- Host based Traffic features: These features examine the connections which established in the past two seconds and which have the same destination host as the current connections.

The dataset contains 21 different types of attacks which can be categorized into 4 types.

- Denial of Service (DoS): Attacker tries to forbid the legitimate users from utilizing the requested services/resources
- Probe (PRB): Attackers attempt to gain information about the target host.
- Remote to Local (R2L): Attackers do not have the account in victim computer, so they try to get access to the computer.
- User to Root (U2R): Unauthorized user tries to gain access to local super user(root) through the network.

III. Traditional Fuzzy C-Means

Fuzzy C-Means clustering algorithm is based on the traditional fuzzy set which was proposed by Bezdek [3]. FCM is the improvement of K-means algorithm which groups the data points based on the membership value. In FCM, the membership of a data point depends on the similarity of the data point to a particular class relative to all other class. Let $X = \{x_1,....x_j,...x_n\}$ be the N data points and $V = \{v_1,....v_i,...v_j\}$ be the set of c centroids. FCM partitions X into c clusters by minimizing the objective function in (1).

$$J = \sum_{j=1}^{N}\sum_{i=1}^{c} u_{ij}^{m} \left\| x_j - v_i \right\|^2 \tag{1}$$

where u_{ij} is the degree of membership of x_j in the cluster i, v_i is the ith cluster center, $\|\cdot\|$ is a distance metric and m is a constant which controls the fuzziness of the resulting partition. Using the Lagrangian method Bezdek derived two necessary condition for minimizing the objective function J as follows:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\left\|\frac{x_j - v_i}{x_j - v_k}\right\|\right)^{\frac{1}{m-1}}} \tag{2}$$

$$v_i = \frac{\sum_{j=1}^{N} u_{ij}^{m} x_j}{\sum_{j=1}^{N} u_{ij}^{m}} \tag{3}$$

The Clustering process begins by randomly choosing the c cluster centers. Further, the membership are calculated based on the relative distance of data point x_j to the centroid v_i using equation (2). The data points which are close to centroids are assigned highest membership value, where as data points far from the centroids are assigned low membership value. After computing the membership of all the data points, the cluster centers are updated using equation (3). The clustering process stops when the successive difference of objective function is less than the pre defined threshold value (ε).

IV. Proposed Method

The technique of fuzzy clustering has become very important in the application of intrusion detection. This is due to the large role of uncertainty nature of an attack. Motivated by this, in this paper we proposed fuzzy clustering based anomaly intrusion detection method. The proposed method consists of three steps: Pre-processing, Feature Selection and Clustering. Fig. 1 shows the block diagram of the proposed system.

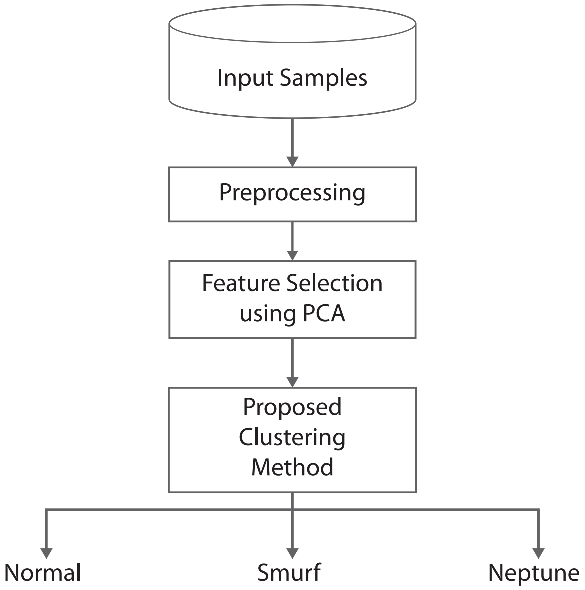

Fig. 1. Block diagram of the Proposed System.

A. Pre-processing

The pre-processing step is performed to make the dataset convenient to be use in the clustering step. The pre-processing will have great impact on the intrusion detection efficiency. In this step we performed two pre-processing steps that are: removing duplicate samples and filling missing values. Dataset contains a large number of duplicate samples. If the dataset contains duplicate samples, then clustering takes more time and also gives inefficient results. To achieve accurate results we removed duplicate samples from the dataset. To fill the missing values, first we divide the whole dataset according to their class and compute the mean value for each feature. Further the missing value is replaced with corresponding feature's mean value.

B. Feature Selection

In this paper, we employed Principal Component Analysis (PCA) [12][15] to select the most discriminative features. PCA is a widely used feature selection method and it is based on linear transformation, which maps data from a high dimensional feature space to lower dimension. The first PCA as the highest contribution to the variance in the original dataset and each succeeding components have remaining variance as possible. The primary advantages of PCA are it is simple, non paramteric and it can preserve large percentage of the total variance with only a few components. Thus, these characteristics motivated us to adopt PCA for feature selection.

Let us consider a training network sample $X = \{x_1, x_2, \ldots x_m\}$ and it has n user behaviour features $F = \{f_1, ., ., ,, f_n\}$. So, totally we have $m \times n$ feature matrix. First we compute the mean of the observations X as follows:

$$\mu = \begin{pmatrix} \overline{x}_1 \\ \overline{x}_2 \\ \vdots \\ \vdots \\ \overline{x}_m \end{pmatrix} \tag{4}$$

Next we compute the covariance matrix i.e

$$S = \frac{1}{n}\sum_{i=1}^{n} \Phi_i \Phi_i^{T} \tag{5}$$

where Φ_i is the standard deviation which is computed as indicated in (6):

$$\Phi_i = X - \mu \tag{6}$$

Φ_i^{T} is the transpose of the standard deviation.

Further, eigen value and corresponding eigen vector of the covariance matrix S is computed. Let $\{(\lambda_1, \xi_1), \ldots (\lambda_k, \xi_k) \ldots (\lambda_m, \xi_m)\}$ are m eigen values and corresponding eigen vectors pairs of the covariance matrix. Next, we choose k eigen vectors corresponding to the largest eigen values. Afterwards, we form $m \times k$ matrix D, whose column consist the k eigen vectors. The representation of data by principal component consist of projecting the original data on to the k dimensional subspace h_k such that:

$$Y_i = D^T \Phi_i = D^T (X - \mu) \tag{7}$$

In the next step we employed the Robust Spatial Kernel FCM (RSKFCM) to cluster these reduced feature matrix.

C. Robust Spatial Fuzzy C-Means

Traditional Fuzzy C-Means (FCM) leads to its non robust result mainly due to: not utilizing the neighbourhood information and use of Euclidean distance. To overcome these problems, Kumar and Harish

[2] proposed a Robust Spatial Kernel FCM (RSKFCM) technique. RSKFCM incorporates spatial information to the conventional FCM membership function and uses kernel distance metric. Experimental results reveal that, RSKFCM gives better clustering results than other FCM variants. Thus, inspired by the good performance presented in [2], we applied RSKFCM method to detect network anomalies.

The main aim of the RSKFCM is to minimize the following objective function J :

$$J = \sum_{i=1}^{c}\sum_{j=1}^{N} w_{ij}^{m} \left\| \Phi(x_j) - \Phi(v_i) \right\|^2 \tag{8}$$

Where Φ is an implicit nonlinear map, and

$$\left\|\Phi(x_j) - \Phi(v_i)\right\|^2 = K(x_j, x_j) + K(v_i, v_i) - 2K(x_j, v_i) \tag{9}$$

Where $K(x, y) = \Phi(x)^T \Phi(y)$ is an inner product kernel function. If we adopt the Gaussian RBF kernel i.e $K(x, y) = \exp\left(-\| x - y \|^2 / \sigma^2 \right)$, then $K(x, x) = 1$. The simplified objective function becomes :

$$J = 2\sum_{i=1}^{c}\sum_{j=1}^{N} w_{ij}^{m}\left(1 - K\left(x_j, v_i\right)\right) \tag{10}$$

Where w_{ij} is the new membership function which combines traditional FCM membership function and neighbourhood membership function. The new membership value is computed as:

$$w_{ij} = \frac{u_{ij}^{p} s_{ij}^{q}}{\sum_{k=1}^{c} u_{kj}^{p} s_{kj}^{q}} \tag{11}$$

Where u_{ij} is the Kernel FCM (KFCM) membership function and s_{ij} is the neighbourhood membership function. Kernel FCM is variant of the FCM, unlike in FCM it uses kernel function as distance metric. The membership function of KFCM is calculated as:

$$u_{ij} = \frac{\left(1 - K\left(x_j, v_i\right)\right)^{-1/(m-1)}}{\sum_{k=1}^{c}\left(1 - K\left(x_j, v_k\right)\right)^{-1/(m-1)}} \tag{12}$$

To compute the neighbourhood membership function, we calculated distance from each samples to other samples and considered k nearest samples. The neighbourhood memebrship function is defined as follows:

$$s_{ij} = \sum_{k \in NK(x_j)} u_{ik} \tag{13}$$

where $NK(x_j)$ represents a array of k nearest samples from x_j . This spatial function represents the probability of sample x_j belongs to i^{th} cluster. In new membership function *p* and *q* parameters controls the relative importance of both functions.

The RSKFCM algorithm is carried out through an iterative optimization of the objective function shown in equation (10) with the update of membership value and cluster centers. The cluster centers are updated using equation (14):

$$v_i = \frac{\sum_{j=1}^{N} w_{ij}^{m} K\left(x_j, v_i\right) x_j}{\sum_{j=1}^{N} w_{ij}^{m} K\left(x_j, v_i\right)} \tag{14}$$

The clustering process stops when the successive difference of the objective function is less than the pre defined threshold value (ε). The individual stages of Robust Spatial Kernel Fuzzy C-Means (RSKFCM) are described in Algorithm 1.

Algorithm 1: RSKFCM Clustering Method

```
Input: Intrusion Samples
Output:  Label
Initialize cluster centers, $\varepsilon$ , m
Repeat
{
    Step 1 Compute all membership values $u_{ij}$ of each sample against centers as:
        $u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\left\|\frac{x_j - v_i}{x_j - v_k}\right\|\right)^{\frac{1}{m-1}}}$    (15)
    Step 2 Compute the new membership value $w_{ij}$ using equation (11)
    Step 3 Calculate the objective function J as follows:
        $J = 2\sum_{i=1}^{c}\sum_{j=1}^{N} w_{ij}^{m}\left(1 - K\left(x_j, v_i\right)\right)$    (16)
    Step 4 Calculate new cluster center values $v_i$ according to expression in (17).
        $v_i = \frac{\sum_{j=1}^{N} w_{ij}^{m} K\left(x_j, v_i\right) x_j}{\sum_{j=1}^{N} w_{ij}^{m} K\left(x_j, v_i\right)}$    (17)
}
Until $\left\{J(i) - J(i-1)\right\} < \varepsilon$
```

V. Experimental Results

To evaluate the proposed method, we conducted experiments on EDA dataset [5]. This dataset is the variant of original KDD dataset. Since, in original KDD dataset smurf, neptune and normal traffic represents 99.3% of the total samples, EDA dataset contains only these three classes.

To handle categorical features, this dataset adds one dummy variable per category into the original KDD dataset, which in turn increases the feature size into 122.

The performance of the proposed method is evaluated using four cluster validity indices, accuracy and false positive rate. For all algorithms in comparison, we set the fuzzy co-efficient m to widely used value 2. All the cluster centers are initialized randomly. We set stopping criteria ε = 0.0001. We implemented and simulated all the algorithms with matlab2013.

A. Evaluation using Cluster Validity Indices

In this section we evalated performance of the propsoed method using four cluster validity indices. These cluster validity indices help to validate whether clustering method accurately presents the structure of the data set or not. Wide varieties of cluster validity indices are proposed in the literature. In this paper we have used four widely used cluster validity functions, namely Partition Coefficient (V_{pc}), Partition Entropy (V_{pe}), Fukuyama-Sugeno function (V_{fs}), and Xie-Beni function (V_{xb}).

Bezdek [20][22] proposed Partition Coefficient (V_{pc}) and Partition Entropy (V_{pe}) which uses only the membership values to evaluate the -cluster validity as indicated in (18) and (19):

$$V_{pc}(U)=\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{c}u_{ij}^{m} \qquad (18)$$

$$V_{pe}(U)=\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{c}u_{ij}^{m}\log u_{ij} \qquad (19)$$

The value of V_{pc} varies between $[\frac{1}{c},1]$ where c indicates the number of clusters. The value of V_{pe} ranges between $[0,\log_a c]$ where c is the number of cluster and a is the base of the logarithm. When V_{pc} is maximal or V_{pe} is minimal, the optimal clusters are achieved.

The Fukuyama-Sugeno function (V_{fs}) [8] is given by:

$$V_{fs}(U,V;X)=\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^{m}\left(\left\|x_j-v_i\right\|^2-\left\|v_i-\bar{v}\right\|^2\right) \qquad (20)$$

where $\bar{v}=\frac{1}{c}\sum_{i=1}^{c}v_i$. V_{fs} uses both the membership information and input data. When V_{fs} value is minimum, the better clustering results are achieved.

The Xie-Beni function (V_{xb}) function, which was initially proposed by Xie-Beni (XB) in [30] and modified by Pal and Bezdek in [18], is defined as indicated in equation (21):

$$V_{xb}(U)=\frac{\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^{m}\left\|x_j-v_i\right\|^2}{n\left(\min_{i\neq k}\left\{\left\|v_i-v_k\right\|^2\right\}\right)} \qquad (21)$$

In V_{xb} the numerator indicates the compactness of the fuzzy partition and denominator indicates the strength of the separation between clusters. When V_{xb} is minimal, the best clustering result is achieved.

To evaluate the performance, we compared our proposed algorithm with traditional Fuzzy C-Means (FCM), Kernel FCM (KFCM) and Spatial FCM (SFCM) methods. Table I presents the performance comparison of the proposed method.

B. Performance Comparison with State-of-the-art Methods

We compared our proposed method with six unsupervised anomaly detection methods. The methods used in comparison are as follows: K-Means [28], Improved K-Means [28], K-Medoids [28], Expectation Maximization [28], Fuzzy C-Means [6], Fuzzy Rough Clustering [6]. We used accuracy and False positive rate as evaluation metrics. Fig. 2 shows the comparison of the proposed method with other methods using accuracy. Fig. 3 shows the comparison of the proposed method with other methods using False Positive Rate.

TABLE I
Performance Comparison

Method	V_{pc}	V_{pe}	V_{xb} $[1\cdot10^{-3}]$	V_{fs} $[-1\cdot10^{6}]$
FCM	0.826	0.201	59.349	2.163
KFCM	0.847	0.196	53.821	2.263
SFCM	0.891	0.121	51.153	2.281
RSKFCM	**0.915**	**0.108**	**32.142**	**2.346**

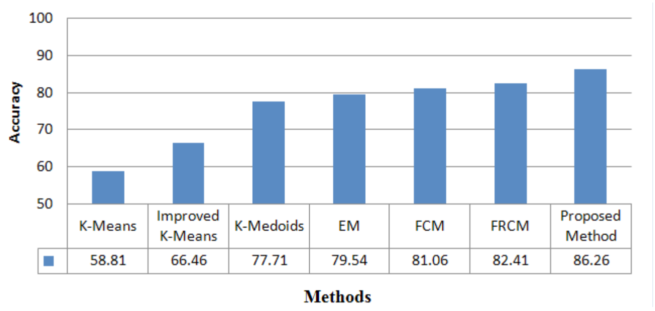

Fig. 2. Comparison of Accuracy.

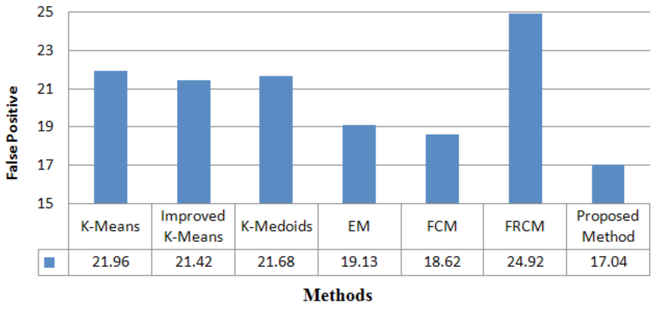

Fig. 3. Comparison of False Positive Rate.

C. Discussion

In Table 1, Fig. 2 and Fig. 3 we can observe that our proposed method outperforms all other methods. As mentioned earlier, our proposed method uses neighborhood information and kernel distance metric. The other methods in comparison, cluster a given sample based on membership value or distance value. Whereas our proposed method considers neighborhood membership value along with the membership value of that sample. This neighborhood information increases the accuracy of the proposed method. On the other hand, other methods use Euclidean distance which is very sensitive to noise. Whereas our proposed method uses Gaussian kernel distance metric. This reduces the noise effect and in turn increases the accuracy.

VI. Conclusion

The existing computer security technologies fail to prevent the threats completely. As a result, Intrusion Detection System (IDS) becomes important component in network security. IDS reduces the manpower needed for monitoring and increases the detection efficiency. In this paper, we presented Fuzzy C-Means based intrusion detection system. Principal Component Analysis is employed to select the most discriminate features. Afterwards a Robust Spatial Kernel FCM is used to cluster the network samples. To evaluate the efficiency of the proposed method, we conducted experiments on standard dataset and compared the results with variants of traditional Fuzzy C-Means methods and other clustering methods. The results inferred that the proposed method outperforms the other methods. The advantage of the proposed clustering method is it considers the neighbourhood membership value and uses kernel distance metric which increases the clustering accuracy and reduces the noise effect. However, the performance of RSKFCM algorithm depends on the cluster center initialization. In future work, Evolutionary algorithm can be used to initialize the cluster centers.

References

[1] Ahmed, Mohiuddin, Abdun Naser Mahmood, and Jiankun Hu. “A survey of network anomaly detection techniques,” *Journal of Network and Computer Applications, vol.* 60, pp. 19-31, 2016.

[2] S. V. Aruna Kumar and B. S. Harish, “Segmenting MRI Brain Images using novel Robust Spatial Kernel FCM (RSKFCM),” *Eighth International Conference on Image and Signal Processing*, pp. 38–44, 2014.

[3] J. C. Bezdek, R. Ehrlich, and W. Full, “FCM: The Fuzzy C-Means clustering algorithm,” *Computers & Geosciences*, vol. 10, no. 2-3, pp. 191–203, 1984.

[4] K. Bharti, S. Shukla, and S. Jain, “Intrusion detection using unsupervised learning,” *International Journal on Computer Science and Engineering*, vol. 1, no. 2, pp. 1865–1870, 2010.

[5] J. Camacho, “Visualizing big data with compressed score plots: approach and research challenges,” *Chemometrics and Intelligent Laboratory Systems*, vol. 135, pp. 110–125, 2014.

[6] W. Chimphlee, A.H. Abdullah, M.N.M Sap, S. Srinoy and S. Chimphlee, Anomaly-based intrusion detection using fuzzy rough clustering. IEEE *International Conference on Hybrid Information Technology, ICHIT’06.* vol. 1, pp. 329-334, 2006.

[7] D. E. Denning, “An intrusion-detection model,” *IEEE Transactions on software engineering*, no. 2, pp. 222–232, 1987.

[8] Y. Fukuyama and M. Sugeno, “A new method of choosing the number of clusters for fuzzy c-means method,” *In Proceedings of Fifth Fuzzy Systems Symp*, pp. 247–250, 1989.

[9] S. Ganapathy, K. Kulothungan, P. Yogesh, and A. Kannan, “A no accuracy weighted fuzzy c–means clustering based on immune genetic algorithm or intrusion detection,” *Procedia Engineering*, vol. 38, pp. 1750–1757, 2012.

[10] F. Guorui, Z. Xinguo, and W. Jian, “Intrusion detection based on the semi-supervised fuzzy c-means clustering algorithm,” in *Consumer Electronics, Communications and Networks (CECNet), 2012 2nd International Conference on* IEEE, pp. 2667–2670, 2012.

[11] S. M. Hameed, S. Saad, and M. F. AlAni, “An extended modified fuzzy possibilistic c-means clustering algorithm for intrusion detection,” *Lecture Notes on Software Engineering*, vol. 1, no. 3, p. 273, 2013.

[12] H. Hotelling, “Analysis of a complex of statistical variables into principal components.” *Journal of educational psychology*, vol. 24, no. 6, pp. 417, 1933.

[13] M. Jianliang, S. Haikun, and B. Ling, “The application on intrusion detection based on k-means cluster algorithm,” in *Information Technology and Applications, 2009. IFITA’09. International Forum on*, vol. 1. pp. 150–152, 2009.

[14] Y. Jun-lan, “Intrusion detection technology and its future trend,” *Information Technology*, vol. 4, pp. 172–176, 2006.

[15] P. Kushwaha and R.. Welekar. “Feature Selection for Image Retrieval based on Genetic Algorithm.” *International Journal of Interactive Multimedia and Artificial Intelligence*, Vol.4, pp. 16- 21.2016.

[16] G. Karthik and A. Nagappan, “Intrusion detection system using kernel fcm clustering and bayesian neural network,” *International Journal of Computer Science and* Information *Technology and Security*, vol. 3, no. 6, pp. 391–399, 2013.

[17] S. Khazaee and M. S. Rad, “Using fuzzy c-means algorithm for improving intrusion detection performance,” in *2013 13th Iranian Conference on Fuzzy Systems*, pp. 27–29, 2013.

[18] S. Khazaee and K. Faez, “A novel classification method using hybridization of fuzzy clustering and neural networks for intrusion detection,” *International Journal of Modern Education and Computer Science*, vol. 6, no. 11, p. 11, 2014.

[19] G. R. Kumar, N. Mangathayaru, and G. Narsimha, “An approach for intrusion detection using fuzzy feature clustering,” IEEE *International Conference on Engineering & MIS (ICEMIS)*, pp. 1–8, 2016.

[20] N. R. Pal and J. C. Bezdek, “On cluster validity for the fuzzy c-means model,” *IEEE Transactions on Fuzzy systems*, vol. 3, no. 3, pp. 370–379, 1995.

[21] N. Pandeeswari and G. Kumar, “Anomaly detection system in cloud environment using fuzzy clustering based ANN,” *Mobile Networks and Applications*, vol. 21, no. 3, pp. 494–505, 2016.

[22] J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media, 2013.

[23] W. Ren, J. Cao, and X. Wu, “Application of network intrusion detection based on fuzzy c-means clustering algorithm,” *Third International Symposium on Intelligent Information Technology Application, 2009. IITA*, vol. 3. IEEE, 2009, pp. 19–22. 2009.

[24] Z. Rustam and A. S. Talita, “Fuzzy kernel c-means algorithm for intrusion detection systems,” *Journal of Theoretical and Applied Information Technology*, vol. 81, no. 1, p. 161, 2015.

[25] R. Sampat and S. Sonawani, “Network intrusion detection using dynamic fuzzy c means clustering.” *Network*, vol. 2, no. 4, 2015.

[26] H Seth, and S. D. Bay. “The UCI KDD Archive [http://kdd. ics. uci.edu]. Irvine, CA: University of California.” Department of Information and Computer Science 152, 1999.

[27] S. Surana, “Intrusion detection using fuzzy clustering and artificial neural network,” Advances in Neural Networks, Fuzzy Systems and *Artificial Intelligence, ISBN-978-960-474-379-7*, 2013.

[28] I. Syarif, A. Prugel-Bennett and G. Wills. Unsupervised clustering approach for network anomaly detection. *Networked Digital Technologies*, pp.135-145. 2012.

[29] H. Wang, Y. Zhang, and D. Li, “Network intrusion detection based on hybrid fuzzy c-mean clustering,” *Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD),* vol.1., pp. 483–486. 2010.

[30] X. L. Xie and G. Beni, “A validity measure for fuzzy clustering,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 13, no. 8, pp. 841–847, 1991.

[31] L. Xie, Y. Wang, L. Chen, and G. Yue, “An anomaly detection method based on fuzzy c-means clustering algorithm,” in *Second International Symposium on Networking and Network Security (ISNNS, 10) Jinggangshan, PR China*, pp. 89–92. 2010.

[32] W. L. Yaseen, A.O Zulaiha and A N M Zakree. “Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system.” *Expert Systems with Applications*, vol.67, pp. 296-303, 2017.

B S Harish

He obtained his B.E in Electronics and Communication (2002), M.Tech in Networking and Internet Engineering (2004) from Visvesvaraya Technological University. He completed his Ph.D. in Computer Science (2011); thesis entitled “Classification of Large Text Data” from University of Mysore. He is presently working as an Associate Professor in the Department of Information Science & Engineering, JSS Science & Technology University, Mysuru. He was invited as a Visiting Researcher to DIBRIS - Department of Informatics, Bio Engineering, Robotics and System Engineering, University of Genova, Italy from May-July 2016. He delivered various technical talks in National and International Conferences. He has invited as a resource person to deliver various technical talks on Data Mining, Image Processing, Pattern Recognition, Soft Computing. He is also serving and served as a reviewer for National, International Conferences and Journals. He has published more than 40 International reputed peer reviewed journals and conferences proceedings. He successfully executed AICTE-RPS project which was sanctioned by AICTE, Government of India. He also served as a secretary, CSI Mysore chapter. He is a Member of IEEE (93068688), Life Member of CSI (09872), Life Member of Institute of Engineers and Life Member of ISTE. His area of interest includes Machine Learning, Text Mining and Computational Intelligence.

S V Aruna Kumar

He received his B.E degree in Computer Science and Engineering from Kuvempu University and M.Tech degree in Software Engineering from Visvesvaraya Technological University, India. He is currently pursuing Ph.D degree in Computer Science from Visvesvaraya Technological University. His area of research includes Machine Learning, Pattern Recognition, Image Processing and Soft Computing.