Abstract: Clustering of attack patterns is used in intrusion detection and discussed in this paper. Fuzzy C-Means (FCM) is a popular method for clustering; the performance of FCM is depends to its parameters radically. In this paper, performance of FCM is evaluated with some Validity Indices and represented in a multi-objective optimization problem. Multi-Objective Simulated Annealing (AMOSA) has high power to solve multi-objective optimization problems, so it is used in this paper to tune the parameters of FCM.

Key-Words:- Intrusion Detection, Clustering, FCM, Multi-Objective Optimization, AMOSA

1 Introduction
Intrusion is the act of wrongfully entering upon, seizing, or taking possession of the property of another. In Information Technology (IT), intrusion is defined as an unauthorized access to or a malicious activity on a computer or an information system [4]. Intrusion detection provides the real-time protection for internal attacks, external attacks and disoperation [1]. Intrusion detection technologies can be divided into two categories: misuse detection and anomaly detection. Misuse intrusion detection firstly groups attack patterns, then manually set up the corresponding detection rules and patterns to construct the intrusion detection system. Anomaly intrusion detection is mainly dependent on the intuition and experience to select the statistic feature to construct the intrusion detection system. Since diversion in network is abruptly, misuse intrusion detection is used more usable. Firstly, there is need to powerful approach to group attack patterns. Since there is no pre-knowledge about structures of attack patterns, clustering is more useable.

The work in this paper has been supported by National Natural Science Foundation of China (61272500), National High-tech R&D Program (863 Program) (2015AA017204) and Beijing Natural Science Foundation (4142008).
2 Related Works

To evaluate the intrusion detection system, two major indicators of performance were of interest [13]: the detection rate and false positive rate. The detection rate (DR) is defined as the number of intrusion instances detected by the system divided by the total number of intrusion instances present in the test set. The false positive rate (FPR) is defined as the total number of normal instances that were (incorrectly) classified as intrusions by the total number of normal instances.

With the development of Internet, the quantitative risk is assessed and automatic decision-making of network security has been especially important. Clustering algorithm is used to mine network security evaluation rules which are utilized to build an algorithm model of network security model to assess decision-making [14]. First of all, network security is based on two basic assumptions. User behavior and procedures are visible and normal behavior and acts of intrusion can be distinguish essentially. Under normal network environment behavior is the mainstream and the invaded is individual. The testing data set may be considered to divide into the normal behavior of the cluster groups and various non-normal mode or small group of cluster. Namely, only are the normal behavior patterns classified into a concentrated clustering set and various type of invaded behavior is not further clustered. Therefore, a normal pattern of cluster groups are established which keeps the normal behavior patterns of behavior away from the invaded as soon as possible.

The proposed algorithm in [15] can speed up the overall computation time and reduce the total number of calculation. The main idea of the proposed algorithm is to refine the initial cluster centers. The algorithm does not note to the value of fuzzification, so the shape of the clusters are not optimize.

Complex network offers a bran-new view for studying complexity [16]. Degree $k_i$ of node $i$ is defined as number of other odes connected by the node. Aggregation coefficient of node $i$ is defined as follow:

$$ C_i = \frac{2E_i}{(k_i(k_i-1))} $$

In characteristic view the formula for aggregation coefficient is changed to

$$ C_i = \frac{\text{the number of triangles connected to node } i}{\text{the number pf triples connected to node } i} $$

The weights between nodes would influence to difference between the networks. In this study of network security evaluation model, whether the nodes are connected must be not only considered, but also the risk conductivity should be considered in the transmission of the risk between two nodes.

3 FCM

Fuzzy C-Means (FCM) is proposed by Bezdek [2]. FCM clustering algorithm determines whether the thing belongs to a certain category on the basis of its membership. Let $X = \{X_1, X_2, \ldots, X_n\}$ is data set and each of data vectors belong to $d$-dimensional space i.e. $X_j \in \mathbb{R}^d$. FCM algorithm minimizes the following objective function with respect to membership degree of each data and cluster centroid [7].

$$ J_m = \sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{ij})^m d^2(X_j, V_i) $$

Such That: $\sum_{i=1}^{c} \mu_{ij} = 1$

Where

$$ d^2(X_j, V_i) = (X_j - V_i)^T A (X_j - V_i) $$

$A$ is a definite matrix, $c$ number of clusters and $V_i$ is centroid of each cluster, $n$ is number of data vectors (or data points), $\mu_{ij}$ membership of specific data vector in specific cluster and $m > 1$ is fuzzification.

FCM is an iterative algorithm. In the first iteration, values of cluster centroids are selected randomly, but in the next iterations they are determined according to the following formula:

$$ V_i = \frac{\sum_{j=1}^{n} (\mu_{ij})^m X_j}{\sum_{j=1}^{n} (\mu_{ij})^m} $$

Membership degrees are recomputed in each epoch.
Iterations of algorithm repeat until the value of $I_{mx}$ do not change.

The probability of selecting $i^{th}$ cluster given the $j^{th}$ data vector is:

$$h(i|X_j) = \frac{1}{\sum_{k=1}^{c} \frac{1}{d^2(X_j, V_k)}}$$

(5)

Fuzzy Covariance Matrix of the $i^{th}$ cluster is calculated with:

$$FCM_i = \frac{\sum_{j=1}^{N} h(i|X_j)(X_j - V_i)(X_j - V_i)^T}{\sum_{j=1}^{N} h(i|X_j)}$$

(6)

There are two parameters influence on performance of FCM: value of fuzzification ($m > 1$) and number of clusters ($n$). To get the best performance of FCM, important parameters of FCM must be optimized simultaneously according to Validity Index. Validity indices are some measures which determine suitable clustering. Validity functions typically suggest finding trade-off between intra-cluster and inter-cluster variability [9]. There are numbers of validity criteria for fuzzy clustering [10, 11], but four of them cover other ones. Bezdek designed Partition Coefficient to measure the amount of “overlap” between clusters [7].

$$F = \frac{1}{n} \sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{ij})^2$$

(7)

Where $n$ is number data vectors and $c$ is number of clusters. Compactness and Separation validity is computed by the following formula:

$$S = \frac{\sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij}^2 \| \mathbf{V}_j - \mathbf{X}_j \|^2}{n \min_{i,j} \| \mathbf{V}_i - \mathbf{V}_j \|^2}$$

(8)

Fuzzy Hyper-Volume [8] is calculated with (9).

$$F_{HV} = \sum_{i=1}^{c} \left[ \det FCM_i \right]^{1/2}$$

(9)

Where $FCM_i$ is (6). Partition Density is:

$$D_P = \frac{A}{F_{HV}}$$

(10)

$$A = \sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij}$$

4 Multi-Objective Optimization

Multi-Objective Optimization (MOO) are necessary when multiple cost functions are considered in the same problem. The aim of MOO is tuning the decision variables to satisfy all objective functions to optimum value. This class of problems is modeled by (11).

$$\text{Optimize } [F_1(X), \ldots, F_p(X)]$$

(11)

Subject to: $g_i(X) \leq 0, h_j(X) = 0$; $i = 1, \ldots, m; j = 1, \ldots, p$

Where $k$ is the number of objective functions, $X$ is the decision vector, $m$ is the number of inequality constraints and $p$ is the number of equality constraints.

This goal causes different between these algorithms and their ancestor Single-Objective Optimization, which is base on concept of best, while the multi-objective optimization uses the concept of dominance. Dominance is defined in [17]:

$$\mathbf{u} = (u_1, \ldots, u_n) < \mathbf{v} = (v_1, \ldots, v_n) \text{ iff } \forall i \in \{1, \ldots, n\}; u_i \leq v_i$$

(12)

In words, a vector $\mathbf{u} = (u_1, \ldots, u_n)$ dominates another vector $\mathbf{v} = (v_1, \ldots, v_n)$ if and only if $\mathbf{u}$ can reach to optimal value for some criteria without causing a simultaneous non-optimal value for at least one criterion. If two vectors cannot dominate each other, they are called as non-dominated vectors.

4.1 Multi-Objective Simulated Annealing

Basic concept in Simulated Annealing is evolution of the solution by simulating the decreasing temperature ($tmp$) in a material, where higher the temperature meaning that higher the modification of the solution at a generation. If temperature of a hot material
decreases very fast its internal structure may divers and material becomes hard and fragile. Decreasing temperature slowly yields higher homogeneity and less fragile material. Evolution of the solution is carried at specific temperature profiles. At the first iterations a diverse set of initial solutions for the problem is produced at higher temperature. And, these solutions are evolved while the temperature decreases to get their local optimums. In multi-objective situation, there are non-dominated solutions which must be kept in the archive, as a candidate of optimal solution.

Along the run of AMOSA algorithm, there are two solutions: current-so and new-so [18]. They can have three states compared to each other: i- current-so dominates new-so, ii- current-so and new-so are non-dominated each other and iii- new-so dominates current-so.

If new-so is dominated by current-so, there may be solutions in archive which dominates new-so. New-so is accepted to the archive by the probability:

\[ p = \frac{1}{1 + \exp(-\Delta)} \]  

(13)

Where \( \Delta \) is differencing between new-so and other solutions which dominated new-so. If there are \( A \) solutions in the archive,

\[ \Delta = \frac{\sum_{i=1}^{A} \Delta_i}{A} \]  

(14)

Solutions can escape from local-optima and reach to the neighborhood of the global-optima by this probable acceptance.

If new-so is dominated by some solutions in the archive, (14) is modified to:

\[ \Delta = \frac{\sum_{i=1}^{A} \Delta_i}{A} \]  

(15)

When new-so is non-dominated with all members in archive, then new-so is set as current-so and it is added to the archive.

If new-so dominates some solutions in the archive, then new-so is set as current-so and it is added to the archive and solutions in the archive which are dominated by new-so are removed.

If new-so is dominated by by some solutions in the archive, then (13) is changed to:

\[ p = \frac{1}{1 + \exp(-\Delta)} \]  

(16)

Where \( \Delta \) is the minimum of the difference between new-so and dominating solutions in the archive. New-so is set as current-so with the probability (13). If new-so is non-dominated by all solutions in the archive it is set as current-so and added to the archive. If new-so dominates some solutions in the archive, it is set as current-so; it is added to the archive; and all dominated solutions are removed from the archive.

5 Proposed Algorithm

The performance of FCM depends on their parameters radically. In other words, attack patterns are clustered very well if we tune the number of clusters and the value of fuzzification. The aim of this paper is optimization of FCM parameters with AMOSA according to validity indices. There are number of validity indices. In [19, 20] some features for suitable clusters are represented. Clusters must have optimum Adherence and Coherence simultaneously. There is need to clusters which have maximum Adherence (means patterns in cluster have the most similarity to each other) and minimum Coherence (means minimum similarity between patterns in different clusters).

Since formulas (7), (8) and (10) represent adherence and coherence of clusters, they must be optimum simultaneously. [18] represents the ability of AMOSA in multi-objective optimization, so AMOSA is used in this paper. Therefore problem is declared as follow:

\[ \text{Optimize } [F_1(x), F_2(x), F_3(x)] \]  

(17)

Where \( F_1(x) \) is Partition Coefficient (7) is, \( F_2(x) \) is Compactness and Separation Validity (8) and \( F_3(x) \) is Partition Density (10).

\( \overrightarrow{x} \) is a vector which has two fields. One field is number of clusters and another field is value of fuzzification. Values of \( \overrightarrow{x} \) are used in FCM, then values of FCM are used in (17). After iterations optimum value for \( \overrightarrow{x} \) includes optimum number of clusters and optimum value of fuzzification can be reached.

In this paper, problem in (17) is tested with a subset of KDD-99 data [21]. KDD-99 is the abbreviation of International Conference on
Knowledge Discovery and Data Mining 1999. KDD-99 data includes a wide variety of intrusions simulated in a network environment. There is need to measures to evaluate the proposed algorithm. 

\[ \bar{x} = (5.1392) \] is an individual, where 5 in number of clusters and 1.392 is value of fuzzification. Thus, according to (7), (8) and (10) respectively:

\[ F_1 = 13.7525, F_2 = 9.6428, F_3 = 22.351 \]

AMOSA, same as other evolution optimization algorithms, starts with a random initial individual. We use \((5,1.392)\) as initial. AMOSA runs for 10 epochs. Table 1, represents the values of \(\bar{x}\) in each iteration.

<table>
<thead>
<tr>
<th>Number of Clusters</th>
<th>Fuzzification</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1.392</td>
</tr>
<tr>
<td>7</td>
<td>1.6465</td>
</tr>
<tr>
<td>10</td>
<td>1.7302</td>
</tr>
<tr>
<td>11</td>
<td>1.7416</td>
</tr>
<tr>
<td>11</td>
<td>1.9384</td>
</tr>
<tr>
<td>12</td>
<td>1.9506</td>
</tr>
<tr>
<td>15</td>
<td>2.001</td>
</tr>
<tr>
<td>14</td>
<td>2.1025</td>
</tr>
<tr>
<td>15</td>
<td>2.1104</td>
</tr>
<tr>
<td>16</td>
<td>2.2138</td>
</tr>
</tbody>
</table>

Table 1. results in each iteration

Since the values in initialization step are set as random, the results of algorithm may change for different runs. Thus, there is need an approach to measure the usability of algorithm in real environment.

A common measure to evaluate the systems is K-Fold [22, 23]. K-Fold technique ensures the quality of simulated system. In other words, K-Fold ensures that the proposed algorithm can be worked in real systems. In K-Fold approach, the set of data divided into K subsets. Approach has K stages. In each stage of approach, K-1 subsets are used as test data and the rest is used as train data. For example, if K=5; 5-Fold approach has 5 stages. In the first stage, subsets: 1, 2, 3, 4 are used as test data and subset 5 is used train data. In the second stage, subsets: 1, 2, 3, 5 are used as test data and subset 4 is used as train data and so on. Each stage has specific Hit Ratio. Average of Hit Ratio represents the power of proposed algorithm in real environment. Hit ratios of 5-fold are 93%, 88%, 87%, 91%, 88%. Therefore, average hit ration is 89.4%.

6 Conclusion

Misuse intrusion detection approaches need to grouping the attack patterns. Because there is no pre-knowledge about patterns, clustering methods are used. Fuzzy C-Means (FCM) is a popular method for clustering. Actually FCM needs to tune the parameters. Tuning the parameters of FCM deeply influence on its performance. There are Validity Indices, which measure the clustering method. Motivation of this paper is tuning of FCM parameters according to partition coefficient, compactness and separation validity and partition density. Validity indices can be seen as objectives, so problem a multi-objective optimization problem. Multi-objective simulated annealing (AMOSA) is used to solve the recent problem. Results of 5-fold cross validation approach ensure that proposed system can be adapt to real environment.

References


[5]. Peiyu LIU, Linshan DUAN, Xuezhi chi, Zhenfang ZHU, “An Improved Fuzzy C-Means Clustering Algorithm Based on Simulated
Annealing”, 2013 10th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)
[10]. M. Halkidi, Y. Batistakis, M. Vazigiannis, “Clustering algorithms and validity measures”, 0-7695-1218-6/01 $10.00 0 2001 IEEE
[15]. Ming-Chuan Hung, Don-Lin Yang, “An Efficient Fuzzy C-Means Clustering Algorithm”, 0-7695-1 119-8/01 $17.00 0 2001 IEEE
[19]. R. Babuska, H.B. verbuggen, “Constructing Fuzzy Model By Product Space Clustering”, Delft University of Technology, P.O. Box 5031, NL_2600 GA Delft, The Netherlands

ISSN: 2367-8995 61 Volume 1, 2016