Routing Attacks Detection Method of Wireless Sensor Network

Yongzhong Li¹, Miao Du² and Yi Li³

ABSTRACT

Security is a critical challenge for creating robust and reliable sensor networks. For the particularity and security threats of wireless sensor networks, we proposed an anomaly detection method based on particle swarm optimization K-means clustering algorithm to detect routing attacks caused by abnormal flows in this paper. K-means clustering algorithm is an effective method has been proved for apply to the intrusion detection system, but it is a part of the optimal solution, rather than the overall optimal solution. Particle swarm optimization (PSO) algorithm which is evolutionary computation technology based on swarm intelligence has good global search ability. With the deficiency of global search ability for K-means clustering algorithm, K-means clustering algorithm based on particle swarm optimization (PSO-KM) is proposed in this paper. This algorithm could overcome falling into local minima and has relatively good overall convergence. Experiments on data sets KDD CUP 99 validate the effectiveness of the proposed method and show that the method has higher detection rate and lower false detection rate.

Keywords: Wireless sensor networks; PSO; K-means clustering; Intrusion detection system

INTRODUCTION

Wireless sensor network (WSN), as a new generation wireless network, can complete large-scale environmental monitoring and target tracking tasks, and have a wide application prospect in scientific, medical, commercial, defense and other fields. Compared with traditional wireless communication networks, wireless sensor networks have its own distinct characteristics, such as strictly limited node resources, intensive distribution, open communication medium and lack of clear boundaries, network defense

¹School of Computer Science, Jiangsu University of Science and Technology, Zhenjiang, China (Email: liyongzhong61@163.com)
²School of Computer Science, Jiangsu University of Science and Technology, Zhenjiang, China (Email: dumiao0118@163.com)
³School of Computer Science, Jiangsu University of Science and Technology, Zhenjiang, China
for example. These characteristics make wireless sensor networks invaded more easily than traditional wireless networks. Therefore, it is particularly necessary to study the anomaly intrusion detection technology of its security and defense mechanism[1]. Intrusion detection, as an active depth protection technology, has become an important part of wireless sensor network safety.

At present, some scholars have applied machine learning methods into wireless sensor network intrusion detection research. Intrusion detection is divided into two main categories: anomaly detection and misuse detection. A normal detection does not need prior knowledge about intrusion and can detect new attacks, so it gets much attention of researchers who work about intrusion detection. The key of abnormal detection is how to establish normal behavior model of the system and how to use this model to determine and judge the abnormal behavior of the system. Anomaly detection for clustering is anomaly detection technology which does not need guidance. It divides similar data into the same cluster and divides dissimilar data into different cluster. It can automatically detect the unknown attacks. PSO [1-3] algorithm converge rapidly, it is easy to adjust parameter and can be applied to the condition when there is large number of samples and dimension of samples is large. K-means clustering algorithm is an effective method has been proved for apply to the intrusion detection system [1-9]. But it is part of the optimal solution, rather than the overall optimal solution. K-mean clustering algorithm usually select several data from the data object randomly as the initial cluster centers, this makes the clustering results generate great uncertainty. How to choose the initial cluster centers becomes important factor which affect the final clustering results.

Because K-Means clustering algorithm depends on the initial value selection and it is easy to converge at the local extreme, this paper combines particle swarm optimization algorithm with K-Means clustering algorithm and proposes K-means clustering Algorithm based on particle swarm optimization(PSO-KM) for intrusion detection. This algorithm has good overall convergence and can effectively overcome the shortcoming which traditional K-Means clustering algorithm is easy to converge at the local minima. Experiment results verify the effectiveness of K - Means clustering algorithm based on particle swarm optimization. Cluster results of PSO-KM algorithm are not under the influence of initial cluster centers. Its global search ability is better than the ability of K-Means clustering algorithm [2-5].

K-MEANS CLUSTERING ALGORITHM BASED ON PARTICLE SWARM OPTIMIZATION

Particle Swarm Optimization (PSO)

Particle swarm optimization algorithm [2-5] is proposed by American social psychology James Kennedy and Russell Eberhart in 1995. The algorithm simulate behavior of birds flying foraging, search for optimal solutions through cooperation between
individuals. It uses the idea that biotic population share information, its concept is simple and is easy to implement, while there are deep intelligence background.

In PSO, each particle is a point of N-dimensional solution space and has a speed (N-dimensional vector). Different particle has individual fitness associated with objective function. Each particle adjusts their flight path according to its flying experience and flying experience of group and move closer to optimal point. The position of i-particle is denoted as: \( X_i = (x_{i1}, x_{i2}, \ldots, x_{in}) \) Flight speed is denoted as: \( V_i = (v_{i1}, v_{i2}, \ldots, v_{in}) \), the best position which i-particle passed is denoted as: \( P_i = (p_{i1}, p_{i2}, \ldots, p_{in}) \) the groups’ best position which it can get is denoted as: \( G = (g_1, g_2, \ldots, g_n) \). In each step, according to PSO algorithm formula which is proposed by Kenddy, particles update their velocity and position according to the following formula:

\[
V_i(t) = w V_i(t-1) + c_1 r_1 (P_i - X_i(t-1)) + c_2 r_2 (G - X_i(t-1))
\]  

\[
X_i(t) = X_i(t-1) + V_i(t-1)
\]

(1)  

(2)

C1 and C2 denote accelerating factor. According to the experience of PSO algorithm, they are usually set C1-C2 = 2. r1 and r2 are two random number between zero and one, w is called inertia weight. Researchers often use a constant \( V_{\text{max}} \) to limit the speed of particles and improve search results. W plays a role which balance global search ability and local search ability. It is essential for the success of the algorithm. Shi and E berhart study on the effect of the W for optimize performance. They found that the larger the W is, the more easily escape from local minima, and the smaller the W is, the more favorably algorithm converges. Then they present a method which makes inertia weight decrease linearly according to number of iterations. In the beginning algorithm uses large inertia weight, it has a strong overall search capability. The later smaller inertia weight is used and local search ability is improved [3-4]. W is calculated as follows:

\[
w = (w_1 - w_2) \times \frac{\text{Max}_i - i}{\text{Max}_i} + w_2
\]

(3)

W1 and W2 are the initial value and final value of inertia weight. Max_i and i are the maximum number of iterations and the current number of iterations for the algorithm, W reduces from 0.9 to 0.4 with the conduct of iteration.

K-Means Clustering Algorithm

Set pattern samples data set as: \( X = \{X_i, i=1,2,\ldots,n\} \), Where \( X_i \) denotes N-dimensional pattern vector. Clustering problem is to find a division: \( C = \{C1,C2,\ldots,Ck\} \), satisfy the condition:
\[ X = \bigcup_{i=1}^{k} C_i \]  
(4)

\[ C_i \neq \emptyset (i = 1, 2, \ldots, k), C_i \cap C_j = \emptyset (i, j = 1, 2, \ldots, k; i \neq j) \] and make the objective function

\[ J_c = \sum_{j=1}^{k} \sum_{x_i \in C_j} d(X_i, Z_j) \]  
(5)

minimize. Zj denotes the center of cluster j.

d(Xi,Zj) denotes the distance between samples and corresponding cluster center.

Using Euclidean distance:

\[ d(X_i, Z_j) = \|X_i - Z_j\| \]  
(6)

Jc denotes sum of the distance between samples and corresponding cluster center.

**K-Means Clustering Algorithm Based on Particle Swarm Optimization**

Code used in PSO-KM is the encoding mode based on cluster center. The position of each particle is composed of k cluster centers. In addition to position, particles have speed and fitness. Because dimension of the sample vector is N, so the position of particles is k × N dimensional variable, the speed of particles is k × N dimensional variable too. In addition, each particle has fitness. The particles can use the following encoding structure:

\[
\begin{bmatrix}
    z_{11} & z_{12} & \ldots & z_{1N} & v_{11} & v_{12} & \ldots & v_{1N} & f(s_1) \\
    z_{21} & z_{22} & \ldots & z_{2N} & v_{21} & v_{22} & \ldots & v_{2N} & f(s_2) \\
    \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
    z_{k1} & z_{k2} & \ldots & z_{kN} & v_{k1} & v_{k2} & \ldots & v_{kN} & f(s_k)
\end{bmatrix}
\]

When the cluster centers are determined, the division of the clustering is decided by the following nearest neighbor rule. If the Xi satisfies \( \|X_i - Z_m\| = \min_{j=1,2,\ldots,k} \|X_i - Z_j\| \), the Xi belongs to the M category. After using the above rule to divide the sample into the each cluster, Jc can be calculated with equation (4), fitness value can be calculated by the following formula:

\[ j(s_j) = \frac{1}{1 + J_c} \]  
(7)

**K-Means Clustering Algorithm Based on Particle Swarm Optimization is following::**

- Input samples Xi and the number of cluster k, each sample is assigned to a particular class randomly as the initial cluster partition, and calculate cluster centers as the initial particle’s position code, calculate the fitness of particles. Initialize velocity of particles and repeated it N times, generate N Initial Particle Swarm.

- For \( t = 1 \) to Max_Generations;
Process particle swarm $P_{(t-1)}$ with PSO-KM algorithm;

- Encode according to the clustering center of particle, determine cluster partition for corresponding particle with the nearest neighbor rule;

- Calculate the new cluster center according to cluster partition, update the fitness value of particle, replace the original code value;

- EndFor $t$;

- Output the best position coded, determine the cluster center $Z$;

In the course of using PSO-KM algorithm, there will be empty cluster when re-divide new generation of particle swarm. If there is an empty cluster, then randomly remove the pattern vector which is furthest from the cluster center from other non-empty clusters, put the vector into the empty cluster, repeat this process until there is no empty cluster in the cluster partition.

**EXPERIMENT RESULTS AND ANALYSIS**

**Experiment Data**

In order to evaluate application effect of PSO-KM algorithm in intrusion detection, this paper use intrusion detection network test data KDD Cup 1999 data set to do experiment [16]. Data set includes variety of simulated invasion in network environment. Each connected instance contains 41 attributes. Intrusion detection algorithm based on unsupervised clustering often is based on an assumption that the number of normal behavior is far greater than the number of intrusion. In the premise of this assumption, intrusion detection can be judged according to the size of cluster. Large cluster correspond to normal data, small cluster often is invasion. However, some invasion such as DoS and Probing, they often generate a lot of intrusion data. And some types of normal system behavior only generate a small amount of data. These two conditions do not meet the assumption and often require separate treatment. In this paper, use data set which contains only the normal data as training data, then detect data set which contains intrusion data, calculate $d$ which denotes distance between $X_i$ and cluster center. When $d$ is less than threshold, $X_i$ is considered normal data, otherwise it is considered as intrusion data. Because the detection is not carried out according to the size of the cluster, it can detect intrusion when the above assumption is not correct.

In the experiment, use "kdd-train-nor"of kddcup data_10_percent as the training data set which contains 97278 picked normal data. Use "corrected" as the test data, which contains 60593 normal data and 250436 intrusion data [9-13].
A shortcoming of K-Means clustering algorithm is that the test results are still subject to the impact of number of clusters. Value of clusters number K and threshold d directly affects the effect of clustering, but there is no very effective method to determine K. This paper determines K and d by experiment, so can only tentatively select different value of K and d to do experiment. Value of K starts from a smaller value and increases constantly. Value of d starts from a larger value and reduces constantly. Table I show results of two clustering algorithm with different K and D.

<table>
<thead>
<tr>
<th>Cluster Number K</th>
<th>Threshold D</th>
<th>K-means clustering algorithm</th>
<th>PSO-KM algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Known attacks (%)</td>
<td>Unknown attacks (%)</td>
</tr>
<tr>
<td></td>
<td>DR</td>
<td>FTR</td>
<td>DR</td>
</tr>
<tr>
<td>20</td>
<td>74.8</td>
<td>0.93</td>
<td>35.4</td>
</tr>
<tr>
<td>30</td>
<td>78.9</td>
<td>2.35</td>
<td>52.1</td>
</tr>
<tr>
<td>40</td>
<td>89.2</td>
<td>3.67</td>
<td>25.5</td>
</tr>
<tr>
<td>50</td>
<td>67.4</td>
<td>3.09</td>
<td>47.6</td>
</tr>
<tr>
<td>50</td>
<td>67.2</td>
<td>2.67</td>
<td>47.7</td>
</tr>
<tr>
<td>60</td>
<td>67.6</td>
<td>2.58</td>
<td>22.8</td>
</tr>
<tr>
<td>70</td>
<td>65.8</td>
<td>3.47</td>
<td>48.7</td>
</tr>
<tr>
<td>70</td>
<td>66.4</td>
<td>2.96</td>
<td>48.7</td>
</tr>
<tr>
<td>80</td>
<td>65.8</td>
<td>2.87</td>
<td>48.2</td>
</tr>
<tr>
<td>Total</td>
<td>71.5</td>
<td>2.73</td>
<td>41.8</td>
</tr>
</tbody>
</table>

The results of Table I shows that average detection rate which is generated by detecting known attacks with PSO-KM algorithm has reached 74.43%. Comparing to K-Means clustering algorithm, it is improved. The false alarm rate of PSO-KM algorithm is just 1.86%. It is lower than false alarm rate of K-Means clustering algorithm. For unknown attacks, the average detection rate of former has reached 52.1% and false alarm rate of former is just 1.73%, average detection rate and false alarm rate of latter are 41.8% and 2.36%.PSO-KM algorithm is more effective in detecting known and unknown intrusions. It can also be seen from the table that performance of PSO-KM algorithm is better when K is 60 and d is 50.
**Data Preprocessing**

The original data include discrete attribute features and continuous attribute features. It does not process discrete attribute features in the experiment, so that each data sample contains only 38 attributes. For continuous attribute features, different attributes have different measure standard, it will produce issue that large numbers cover up small number. Some attribute features of data will be concealed. In order to solve this problem, attribute feature value of data must be standardized. Transform as follows [4-5]:

Calculate average absolute deviation $S$:

$$ S = \frac{1}{n} \sum_{i=1}^{n} (X_i - m) $$

(8)

$X_i$ denotes data sample, $m$ denotes sample mean,

$$ m = \frac{1}{n} \sum_{i=1}^{n} X_i $$

(9)

Calculate standardized data

$$ Y_i = \frac{X_i - m}{S} $$

(10)

This is equivalent to that attribute features of original instance is mapped to standard attribute space by using statistical characteristics. It is help to reduce the above problems.

**Training and Testing**

Pseudo-code for training and testing the model is as follow:

- For $i=1$ to $n$
- For input samples $X_i$, calculate standardized sample $Y_i$ with formula (5) and formula (6)
- EndFor $i$
- Input sample data set $Y$ and the number of clusters $k$, set the size of particle swarm $M$, initialize population $P (0)$, process data set with PSO-KM algorithm. For test data $X$, calculate the value of standardized $Y$ and calculate $d (X_i, Z_j)$ which denotes the distance between $Y$ and the cluster center
- If $d(Y,Z) < \tau$ // $\tau$ denotes threshold
  - $X$ is Normal data
- Else
Comparison and Analysis of Test Results

Matlab7 is used for programming in experiment. Computer configuration is Intel Pentium 2.80GHz CPU,. We respectively process training data with PSO-KM algorithm and K means clustering algorithm [14-15]. Figure 1 shows the results of computing processing.

![Figure 1. Comparison of two algorithm function.](image)

Thin dotted line is convergence curve of K-means cluster algorithm, and thick solid line is convergence curve of PSO-KM algorithm. From the figure we can know, under the same conditions, convergence speed of K-means cluster algorithm is faster, but it is easy to fall into local minima (between generation ten and generation twenty). Ability of overall optimization of PSO-KM algorithm is better than the ability of K-means cluster algorithm. It has no degradation phenomena of stochastic optimization. Therefore, convergence is relatively stable with faster convergence rate. Figure 2 shows ROC curve of the detection rate (DR) and false detection rate (FR).
Figure 2. Comparison on detection Rate with false Alarm Rate.

The figure shows that it has the best performance when FR is 2.8%. DR is 82% at the same time. It shows that the algorithm is feasible and effective in detecting unknown intrusion.

Table II shows that PSO-KM algorithm has high detection rate for detecting Probing, U2R and R2L. Its detection rate for detecting DoS is relatively low. Because DoS is the most sample in the sample set. Because there is miss test, abnormal data may be labeled as normal data and reduce the detection rate for detecting DoS attack. In conclusion, for PSO-KM algorithm, detection rate for detecting known attacks has reached 75.82%, detection rate for detecting unknown attacks has reached 60.8%. Compared to traditional K-means clustering algorithm, it is better to detect intrusion.

Table II. Comparing the results of two algorithms detection rate.

<table>
<thead>
<tr>
<th>Attack types</th>
<th>K-means clustering algorithm</th>
<th>PSO-KM algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Known attacks (%)</td>
<td>Unknown attacks (%)</td>
</tr>
<tr>
<td>Probing</td>
<td>81.5</td>
<td>7.46</td>
</tr>
<tr>
<td>DOS</td>
<td>29.6</td>
<td>8.84</td>
</tr>
<tr>
<td>U2R</td>
<td>75.6</td>
<td>33.83</td>
</tr>
<tr>
<td>R2L</td>
<td>94.7</td>
<td>65.42</td>
</tr>
</tbody>
</table>
Invasion types of data set can be broadly divided into four categories: Probing, DoS, U2R, R2L by attack means. Table 2 shows comparison of the detection results of four attacks when K is 60 and d is 50.

CONCLUSION

PSO-KM algorithm combines particle swarm optimization algorithm with the traditional K-means clustering algorithm. The clustering results are not under the influence of initial cluster centers. Compared to traditional K-means clustering algorithm, it has better ability of overall optimization. Use particle swarm optimization to guide K-means clustering algorithm to select initial cluster centers, it makes cluster easily converge to overall optimization. It gets rid of that selection of initial cluster centers impacts cluster results. Simulation experiment on data sets KDD CUP 99 shows that PSO-KM algorithm is effective method when dealing with large data sets. Experimental results show that detection rate of PSO-KM is improved for detecting known attacks and unknown attacks. At the same time, false detection rate greatly reduces. It improves application value of K-means clustering algorithm in the field of intrusion detection.

ACKNOWLEDGMENT

This paper is supported by Research fund of University of Jiangsu Province(15KJD520066) and Jiangsu University of Science and Technology’s Basic Research Development Program (No. 2015DX006J)

REFERENCES


