International Journal of Security and Its Applications
Vol.9, No.12 (2015), pp.115-128
http://dx.doi.org/10.14257/ijsia.2015.9.12.12

An Algorithm of Clustering by Density Peaks Using in Anomaly Detection

Chunyong Yin[1], Sun Zhang[1], Zhichao Yin[2] and Jin Wang[1]

[1] *School of Computer and Software, Jiangsu Engineering Center of Network Monitoring, Jiangsu Key Laboratory of Meteorological Observation and Information Processing, Nanjing University of Information Science & Technology, Nanjing 210044, China*

[2] *Nanjing No.1 Middle School, Nanjing, Jiangsu, Postal code 210001, China*

Abstract

With the development of the networks, the security of computer networks is becoming more and more serious. The information openness, sharing and interconnection are three important characteristics of computer networks. However, the amounts of intruders and attackers have been grows with the popularization of computers. Therefore, the focus of network security is preventing systems from being invaded effectively. Intrusion detection as a key technology of network security active defense system is designed to distinguish normal behaviors and attack behaviors. Intrusion detection is divided into misuse detection and anomaly detection, and using clustering algorithm is one of the most effective methods for anomaly detection. In this paper, a clustering algorithm based on fast search and find of density peaks is used to distinguish the normal and abnormal network connections to achieve the purpose of anomaly detection. The performance of the algorithm is tested by a data set selected from KDD CUP99. Experiment results show that this algorithm is more suitable than the traditional K-means in data sets containing a large amount of data and uneven density distribution.

Keywords: *intrusion detection; anomaly detection; clustering; density peaks*

1. Introduction

In April 1980, Anderson issued a report for US Air Force entitled "Computer Security Threat Monitoring and Surveillance" [1]. He first proposed the concept of intrusion detection, and divided this thread into three types, including external penetration, internal penetration and illegal behavior. Then he proposed the theory of using audit tracking data to monitor intrusion activities.

Intrusion detection is the key technology of network security active defense system, which is designed to detect and classify the behaviors of networks and hosts. It checks the networks and systems whether there are behaviors of violating the security policy and signs of being attacked by collecting and analyzing the information of network behavior, security log, audit data and some key points in computer system.

Intrusion detection, as a proactive security technology, provides real-time protection against internal attacks, external attacks and false operation and it can intercept and response to intrusion before the network system is endangered. As a result, it is considered as the second security gate after the firewall to monitor the network without affecting the performance of the network. Intrusion detection is implemented by the following tasks: monitoring, analyzing the activities of user and system, auditing system structure and weaknesses, identifying the modes of attacks and sending warning to the relevant persons, analyzing abnormal behavior patterns, evaluating the integrity of the system and data file, and identifying the behaviors against security policies.

ISSN: 1738-9976 IJSIA

Intrusion detection is an effective supplement of the firewall, which helps the system to deal with the network attacks, extends the security management capabilities of the system (including security audit, surveillance, attack identification and response), and improves the integrity of the information security infrastructure.

In terms of a successful intrusion detection system, it can not only make the system administrators to understand any changes in the network system (including procedures, documents and hardware equipment, *etc*.), but also provide a guide to the development of network security policy. The more important thing is that it should be managed and configured easily so that people who do not have enough related knowledge can easily obtain a system with full security protection. Moreover, the scale of intrusion detection should be changed by the network threat, system structure and security needs. Intrusion detection system need to make a correct response in time after identifying the network have a breach, including cutting off the network connection, recording events and alarms, *etc*.

From different aspects, intrusion detection system can be classified as follows [2]:

(1) According to the detecting methods, it can be divided into anomaly detection and misuse detection. The anomaly detection is using the information of the normal behavior from the monitoring system as the basis of the intrusion detection system. The misuse detection is using the information of attacks that have been known (knowledge, pattern, *etc*.) to detect the intrusion and attack.
(2) According to the different sources of data, it can be divided into the intrusion detection based on hosts and that based on networks.
(3) According to the different response to attacks of the intrusion detection system, the intrusion detection system can be divided into active intrusion detection system (real-time intrusion detection system) and passive intrusion detection system (IDS). The active intrusion detection system can detect the vulnerabilities in the target system, automatically fix the vulnerabilities in the target system, and force suspicious users (possible intruders) to exit the system or close the related services. However, the passive intrusion detection system just warns to system security administrators and administrators decide to use some methods to deal with the attacks.

Anomaly detection is a popular topic in current intrusion detection research field. The main methods are collecting a lot of normal event data to establish the model of normal activities, and calculating the offset of current event data and normal model. The activity will be considered as abnormal activity if the offset exceeds the preset threshold [3]. This paper is using clustering algorithm to realize the purpose of anomaly detection.

2. Application of Clustering Analysis in Network Intrusion Detection

In recent years, data mining [4] is one of the new application technologies and draws much attention. It melds different methods and technologies from different fields, such as artificial intelligence, statistics, database, machine learning and so on. It can use analytical tools to extract some useful and potential information from a large amount of data. People can use this technology to find out interesting information for industry, finance and other fields. In network security, data mining is also a useful method of botnets detection [5, 6]. A new technology, SVM (support vector machine), is used in different fields, such as intrusion detection, botnets detection and so on [7, 8].

Intrusion detection process is the process of using the large amount of collected data, such as host system log, audit records and network data packet, *etc*, and analyzing this data to find intrusions and anomalies. From the aspect of data, intrusion detection process is a process of data analysis. Therefore, we can use data mining to analyze the large amount of data, extract enough adequate and potential information, and abstract an effective model to compare and determine the characteristics of the intrusion detection.

Clustering analysis is an important method in data mining, which has no supervision. It can divide event data into several groups or clusters. The data in each cluster is highly similar to each other. Because of this characteristic, cluster analysis is suitable for anomaly detection. By using proper clustering algorithm, the normal behaviors and abnormal behaviors are clustered into their respective clusters.

Unsupervised anomaly detection based on clustering analysis is one of the main research methods in intrusion detection. This intrusion detection method can detect unknown intrusions quickly from the data. It assumes that the normal data in the network is much more than the abnormal data, and the difference between normal data and abnormal data is obvious. Specific detection process is as follows:

Step 1: The collection of network data. Firstly, monitoring and detecting the network data, using packet capture tools to collect original data (mainly the connection records), and standardizing data.

Step 2: : The standardization of collected data. The collected data often contain a variety of data formats and noises, so it is necessary to make it standardized and transform it into uniform data format which is suitable for analysis. Through this step, it can improve data quality and reduce false positive rate.

Step 3: clustering analysis. After the standardization of data, we get the data sets that can be used for clustering analysis. This step needs to select the appropriate clustering algorithm to analyze, classify the data sets, and distinguish the normal and abnormal records.

Step 4: The optimization of the initial clustering set. In step 3, the initial clusters are obtained, and the normal and abnormal records are classified into different clusters. This step optimizes the initial clustering results, puts records that are similarly with each other into one cluster, and marks the cluster whose data size is larger as normal cluster.

Step 5: real-time intrusion detection. According to the optimization results of step 4, intrusion detection is carried out. And it makes the corresponding processing according to the results, such as alarm processing.

3. Researches on Clustering Algorithm

In the clustering algorithms commonly used, hierarchical clustering algorithm complexity is too high, and may be clustered into a chain. Traditional K-means algorithm needs to set the value of parameter K and determine the clustering centers. The setting of these parameters has a decisive impact on the clustering results. Therefore, the relevant researchers have proposed many improved algorithms.

In paper [9], researchers choose more suitable clustering centers by using an improved K-means algorithm to improve the weighted Euclidean distance formula. The information entropy theory is introduced into the unsupervised clustering algorithm. It proposes a new method of similarity calculation and the selection of the initial clustering center, and this algorithm is applied to anomaly detection. The method first filters the data set to reduce the isolated points in order to reduce the negative impact of these isolated points on clustering results. Second, it uses the algorithm based on the most large distance selection to choose the initial clustering centers, and introduces the method of information entropy to define a property weighted Euclidean distance. This weighted Euclidean distance is used to calculate the similarity in the whole clustering process. Last, the records are divided into different clusters in the iterative process.

Researchers get better results through the combination of simulated annealing algorithm and clustering algorithm in paper [10].The simulated annealing algorithm was first introduced by N. Metropolis in 1953. But until 1983, Kirkpatrick used this algorithm to solve combinatorial optimization problems, and mainly to find the best solution to problems. Simulated annealing algorithm determines whether to accept the transfer from

the current solution to the new solution by the transfer probability relative to Metropolis criteria. At the beginning, the control parameter T has a larger value, and this parameter T will decrease slowly after enough transfers. It repeats this process until meeting a certain stopping criterion. Therefore, simulated annealing algorithm can be considered as the iteration of Metropolis algorithm used to control parameter decreasing. Simulated annealing algorithm is an effective algorithm to solve combinatorial optimization problems. It is used widely in different fields, such as optimal control, machine learning, neural network and other problems of local optimization. K-means clustering algorithm can be considered as the problem of optimization in fact, so simulated annealing algorithm can be used to optimize K-means, and solve the limitation of K-means. In this way, the algorithm can jump out of the local optimal solution and achieve the global optimal solution.

In the traditional K-means algorithm, there are some problems affect the clustering results, including isolated points, the setting of parameters and the selection of initial clustering points. The paper [11] improves the treatment of isolated points and the methods to determine initial clustering points, and researchers apply the improving algorithm in anomaly detection. In this paper, an algorithm of clustering by density peaks is used to select clustering centers just by calculating the distance of each point. By using this algorithm, we cluster normal networks and abnormal networks into different clusters and achieve the purpose of anomaly detection.

4. A Clustering Algorithm using in Anomaly Detection

4.1. Clustering by Fast Search-and-Find of Density Peaks

K-means [12] is a kind of indirect clustering algorithm based on the similarity measure between samples, which belongs to non supervised learning algorithm. This algorithm uses parameter K as the number of clusters, and divides N samples into different clusters, so that the samples in each cluster has a high degree of similarity, and the similarity between each cluster is lower. The similarity is calculated according to the average value of the objects in a cluster which is regarded as the center of the cluster. This algorithm first randomly selects K samples, and each of them represents the center of a cluster. For other sample, they are assigned to the most similar cluster according to the distance between samples and the cluster centers. Then, it calculates the new center of each cluster, and repeats the above process until the criterion function converges. However, the traditional K-means cannot detect the data with non spherical distribution.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a representative clustering algorithm based on density. It defines clusters as the largest collection of the density connected points. It can divide clusters with high density and can cluster arbitrary shape distributions. DBSCAN needs two parameters, the scanning radius (EPS) and the minimum number of points (MINPDS). It begins with choosing an unvisited point arbitrarily and find out all the points within the distance of EPS. If the number of points in the vicinity is not less than MINPDS, this point and all the points nearby form a cluster and are marked as visited. Otherwise, the point is marked as noise point. Then, it handles all the unvisited points to expand the clusters. However, DBSCAN is very sensitive to the parameters users defined, and subtle differences may lead to very different results. The choice of parameters has no rules to follow, and it can only rely on experience to determine.

Rodriguez, Alex and Laio Alessandro propose a new clustering algorithm based on fast search- and-find of density peaks [13]. It is assumed that a clustering center has higher local density than other points around it, and the distance between these points and this clustering center is more recent than the other clustering centers. The main calculation formulas of this algorithm are:

Each point needs to calculate two parameters, ρ and δ. Parameter ρ is the local density, and parameter δ is the minimum distance between this point and other points which have higher local density. The local density is defined as:

$$\rho_i = \sum_j \chi(d_{ij} - d_c) \quad (1)$$

If x<0, $\chi(x)$=1, else $\chi(x)$=0. The parameter dc is a radius to calculate the local density. The paper sort the distances between each point from low to high and choose the distance which is in the position of 20% of the total as the value of dc.

The value of δ_i is defined as:

$$\delta_i = \min_{j:\rho_j > \rho_i}(d_{ij}) \quad (2)$$

As the point which has the maximum local density, δ is the maximum distance between it and other points.

$$\delta_i = \max(d_{ij}) \quad (3)$$

We can select proper points as clustering centers by calculating ρ and δ. Clustering centers should have relatively higher ρ and higher δ. As for isolated points, they have higher δ but lower ρ.

In this algorithm, we do not use a noise-signal cutoff. We find for each cluster a border region, defined as the set of points assigned to that cluster but being within a distance d_c from data points belonging to other clusters, and find the point of highest density within its border region for each cluster. The parameter ρ_b is the density of that point. The point of the cluster whose density is higher than ρ_b are considered part of the cluster core, and the others are considered part of the cluster halo.

4.2. Comparison and Analysis

In this paper, we first use two kinds of classical data sets to compare and analysis this clustering algorithm and traditional K-means. The concrete information of two data sets is shown in Table 1 and Table 2.

Table 1. Aggregation

Label	1	2	3	4	5	6	7
Number	45	170	102	273	34	130	34

Table 2. Spiral

Label	1	2	3
Number	101	105	106

In the experiment of Aggregation, we select the parameter dc from the position of 15% and calculate the local density and the minimum neighbor distance of each point, and use two values to establish two-dimensional coordinate department as Figure 1 shows.

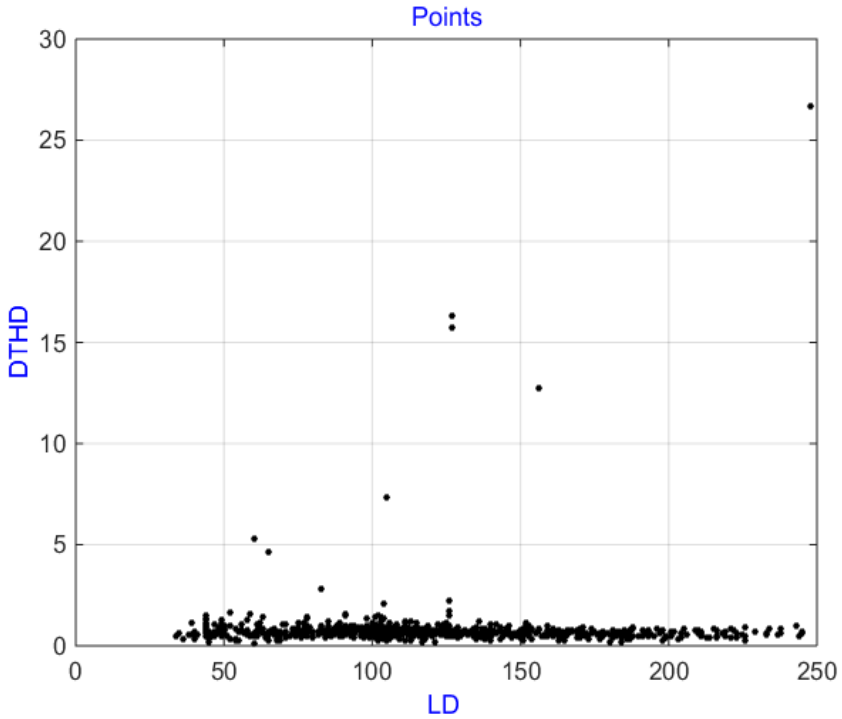


Figure 1. Points of Aggregation

In Figure 1, we can find some points with higher ρ and higher δ and these points may be the clustering centers. Then, we assign other points to these clustering centers. The point in a cluster has the same label as the clustering center, and we compare the label of each point with that before clustering. The results of comparison are shown in Table 3.

Table 3. The Comparison of Aggregation

Label	1	2	3	4	5	6	7
Number	45	158	102	273	34	62	0

In Table 3, we count the number of points that can be correctly labeled. Then we use Aggregation to analyze K-means. The parameter K is 7 and the detailed clustering results is shown in Figure 2.

```
Time taken to build model (full training data) : 0.07 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       66 (  8%)
1      204 ( 26%)
2      101 ( 13%)
3       47 (  6%)
4      168 ( 21%)
5      137 ( 17%)
6       65 (  8%)


Class attribute: class
Classes to Clusters:

   0   1   2   3   4   5   6  <-- assigned to cluster
   0   0   0  45   0   0   0 | 1
   0   0   0   2 168   0   0 | 2
   1   0 101   0   0   0   0 | 3
   0 204   0   0   0  69   0 | 4
   0   0   0   0   0  34   0 | 5
  65   0   0   0   0   0  65 | 6
   0   0   0   0   0  34   0 | 7

Cluster 0 <-- No class
Cluster 1 <-- 4
Cluster 2 <-- 3
Cluster 3 <-- 1
Cluster 4 <-- 2
Cluster 5 <-- 5
Cluster 6 <-- 6

Incorrectly clustered instances :       171.0     21.7005 %
```

Figure 2. First Results of K-Means

Then, we use Spiral to analyze these two algorithms. We select the parameter dc from the position of 50% and calculate the parameters ρ and δ for establishing two-dimensional coordinate department as it is shown in Figure 3.

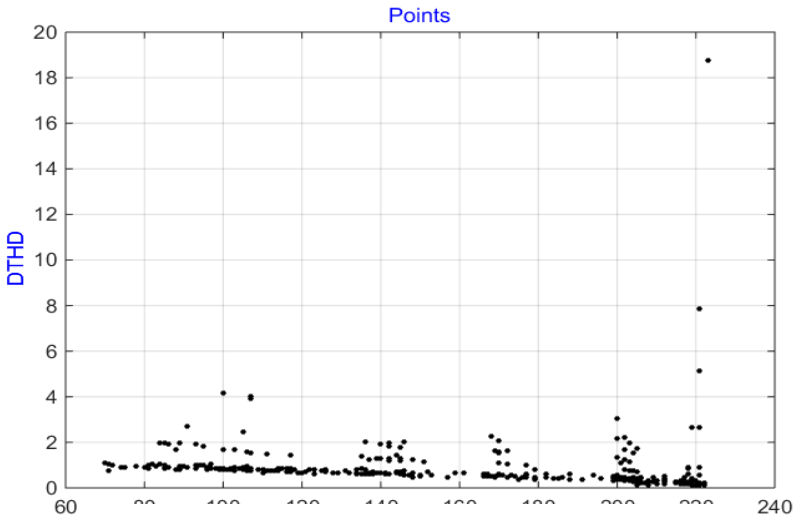


Figure 3. Points of Spiral

From Figure 3, we choose suitable points as clustering centers and assign points to each clustering center. Table 4 is the results of comparing the labels assigned with that before clustering.

Table 4. The Comparison of Spiral

Label	1	2	3
Number	23	51	38

Likewise, we use Spiral to analyze K-means, and the number of clusters is 3. The results of K-means are shown in Figure 4.

```
Time taken to build model (full training data) : 0.02 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       100 ( 32%)
1       106 ( 34%)
2       106 ( 34%)


Class attribute: class
Classes to Clusters:

  0  1  2  <-- assigned to cluster
 33 35 33 | 1
 34 36 35 | 2
 33 35 38 | 3

Cluster 0 <-- 1
Cluster 1 <-- 2
Cluster 2 <-- 3

Incorrectly clustered instances :       205.0    65.7051 %
```

Figure 4. Second Results of K-means

After getting the clustering results of two algorithms, we can calculate accuracy rate. The results are shown in Table 5. Algorithm 1 is the algorithm of clustering by density peaks, and Algorithm 2 is K-means.

Table 5. The Comparison of Two Algorithms

Accuracy Rate	Algorithm 1	Algorithm 2
Aggregation	85.5%	78.3%
Spiral	35.9%	34.3%

From Table 5, we can see that algorithm 1 is more accurate than algorithm 2 for Aggregation. For Spiral, the accuracy of two algorithms is not high, but algorithm 1 is slightly better than algorithms 2.

In the experiment, the algorithm of clustering by density peaks relies on choosing the appropriate value of dc. We first use dc to calculate ρ and δ. Then, we use ρ and δ as the horizontal and vertical coordinates to draw two-dimensional coordinates department, and select appropriate clustering centers.

5. Experiment and Result Analysis

In this experiment, data set directly choose from KDD CUP99. KDD CUP99 consists of 4898431 records, each of which is marked as normal or abnormal, with 3925650 abnormal records, including 22 types of attacks. These 22 attack types can be divided into four categories: DOS, R2L, U2R and Probing. Each record in KDD CUP99 has 41 attributes, and we ignore the first four attributes, and the remaining 37 attributes are used for anomaly detection.

In this paper, we use Accuracy Rate (AR), Detect Rate (DR) and False Positive Rate (FPR) to evaluate the clustering results. AR is the ratio of the amount of records which are correctly labeled and the total amount of data set. DR is the proportion of the total amount of records that are correctly detected as normal. FPR is the proportion of the total amount of records that are falsely detected as abnormal.

5.1. The First Experiment

In the first experiment, 300 records were extracted from KDD CUP99, the number of the class label which is normal, Neptune and back each are 100. The record with a class label of normal is normal data, and the other records are abnormal data. The detailed clustering results of each cluster are shown in Table 6.

Table 6. Clustering Results of Algorithm 1 in the First Experiment

Label	Normal	Neptune	Back
Number	99	75	100

Then, we use this data set to analyze K-means, and the clustering results of K-means are shown in Fig. 5.

After using this data set to analysis the algorithm of clustering by density peaks and K-means, the results are obtained as shown in Table 7.

Table 7. First Analysis Results

Rate	Algorithm 1	Algorithm 2
AR	91.3%	100%
DR	100%	100%
FPR	1%	0

```
=== Model and evaluation on training set ===

Clustered Instances

0       100 ( 33%)
1       100 ( 33%)
2       100 ( 33%)


Class attribute: class
Classes to Clusters:

   0   1   2  <-- assigned to cluster
 100   0   0 | 1
   0 100   0 | 3
   0   0 100 | 4

Cluster 0 <-- 1
Cluster 1 <-- 3
Cluster 2 <-- 4

Incorrectly clustered instances :       0.0       0      %
```

Figure 5. Results of K-means in the First Experiment

5.2. The Second Experiment

In the second experiment, we extract 1472 records from KDD CUP99, the amount of records with normal label is 700. The amount of records with Neptune label is 344, the amount of the other records with back label is 428. The two-dimensional coordinate department composed of the parameters ρ and δ is shown in Fig.6. The detailed clustering results of each cluster are shown in Table 8.

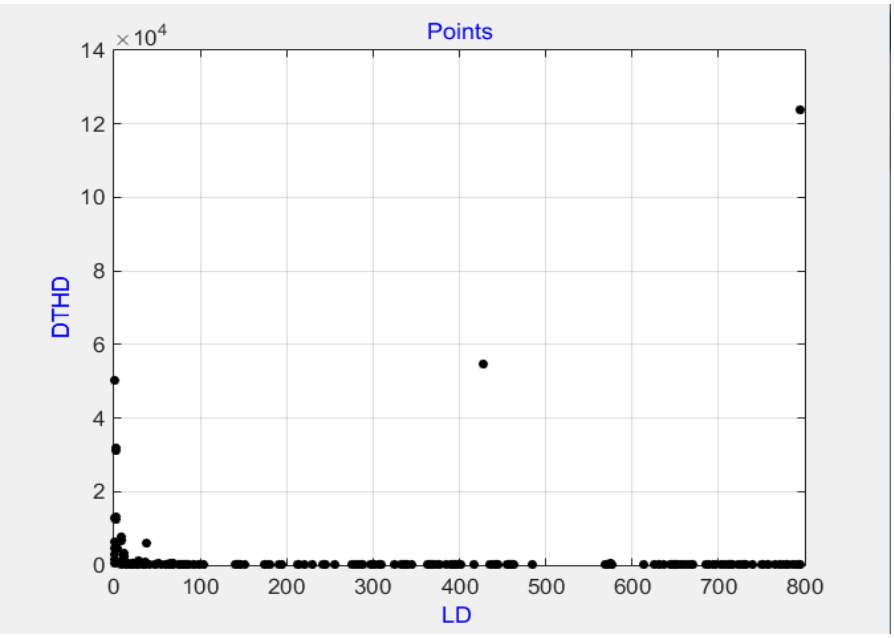


Figure 6. Points of Data Set

Table 8. Clustering Results of Algorithm 1 in the Second Experiment

Label	Normal	Neptune	Back
Number	700	0	428

Then, we use this data set to analyze K-means, and the clustering results of K-means are shown in Figure 7.

After using second data set to analysis these two algorithms, the results are obtained as shown in Table 9.

Table 9. Second Analysis Results

Rate	Algorithm 1	Algorithm 2
AR	76.6%	71.8%
DR	55.4%	55.4%
FPR	0%	0

In Table 7 and Table 9, algorithm 1 is the algorithm of clustering by density peaks, and algorithm 2 is K-means. The number of clustering centers is set to 3. As we can see from Table 7, when the amount of records is 300, algorithm 2 is better than algorithm 1, and algorithm 2 can accurately mark each record, and algorithm 1 will produce false positives.

From Table 9, we can see that, when we increase the amount of data set to 1472, the accuracy rate of algorithm 1 is higher than that of algorithm 2.These two algorithms have the same detection rate, and false positive rate. Algorithm 1 can divide Normal and Back, but algorithm 2 will divide points with the label of Back into two clusters. From Table 5, it can be seen that the accuracy rate of algorithm 1 is higher than that of algorithm 2 for data set with many attributes.

```
Time taken to build model (full training data) : 0.11 seconds

=== Model and evaluation on training set ===

Clustered Instances

0         71 (  5%)
1       1044 ( 71%)
2        357 ( 24%)


Class attribute: class
Classes to Clusters:

   0   1   2  <-- assigned to cluster
   0 700   0 | 1
   0 344   0 | 3
  71   0 357 | 4

Cluster 0 <-- No class
Cluster 1 <-- 1
Cluster 2 <-- 4

Incorrectly clustered instances :       415.0    28.1929 %
```

Figure 7. Results of K-Means in the Second Experiment

6. Conclusions

In this paper, the algorithm of clustering by fast search-and-find of density peaks is used to achieve the purpose of anomaly detection. The main method is using the hosts' normal and abnormal network connection characteristics as the clustering object, and obtained the clustering centers through the set of parameters. This algorithm only considers the distance between each point, and has lower complexity than the traditional K-means.

We use a part of the records from KDD CUP99 to analysis the performance of two algorithms. In two experiments, we use different amounts of data set to test, and we can conclude that the algorithm of clustering by fast search-and-find of density peaks is suitable for the data set with a large number of records and obvious differences. The shortage of the algorithm is that the choice of clustering centers needs to be tried many times, and different clustering centers can lead to different results. How to choose more appropriate clustering centers will be the focus of the next research.

Acknowledgements

This paper is a revised and expanded version of a paper entitled "Anomaly detection based on clustering by density peaks" presented at COMCOMS 2015, Hanoi, Vietnam, October 22-24, 2015. This work was funded by the National Natural Science Foundation of China (61373134, 61402234), and by the Industrial Strategic Technology Development Program (10041740) funded by the Ministry of Trade, Industry and Energy (MOTIE) Korea. It was also supported by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), Jiangsu Key Laboratory of Meteorological Observation and Information Processing (No.KDXS1105) and Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology (CICAEET). Prof. Jin Wang is the corresponding author

.

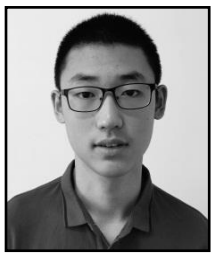
References

[1] J P Anderson. Computer Security Threat Monitoring and Surveillance**(1980),** http://csrc.nist.gov/publications/history/ande80.pdf**.**
[2] Y.R .Zhang, S.P.Xiao and G.Y.Wang, 'An Overview of Intrusion Detection Technology Based on Machine Learning', Computer Engineering and Applications **(2006),** No.02, pp.7-10.
[3] G.Shen, Y.J.Su and X.Liu, 'Anomaly Detection Based on Aggregated Network Behavior Metrics', Computer Engineering and Science **(2010)**, Vol.32, No.03, pp.41-38.
[4] R.F, Erbacher, Walker, K.L, Frincke and D.A , 'Intrusion and misuse detection in large -scale systems', IEEE Computer Graphics and Applications **(2002)**, Vol.22, No.01, pp. 38-47.
[5] C. Yin, 'Towards accurate node-based detection of P2P botnets', Scientific World Journal **(2014)**, 24, June, pp:425491-425491.
[6] C. Yin, M. Zou, D. Iko and Jin Wang, 'Botnet Detection Based on Correlation of Malicious Behaviors', International Journal of Hybrid Information Technology **(2013)**, Vol.6, No.6, pp: 291-300.
[7] B. Gu, S. Victor . Sheng, Keng Yeow Tay, Walter Romano, and Shuo Li, 'Incremental Support Vector Learning for Ordinal Regression', IEEE Transactions on Neural Networks and Learning Systems **(2014)**,Vol.26, No.7, pp:1403-1416.
[8] B. Gu, S Victor. S., Zhijie Wang, Derek Ho, Said Osman and Shuo Li, 'Incremental learning for ν-Support Vector Regression', Neural Networks **(2015)**, Vol.67, pp:140-150.
[9] Z.Chen and G.C. Luo, 'Improved K- Means Algorithm Using in Anomaly Detection', Chongqing Science University of Technology (NATURAL SCIENCE) **(2015)**, Vol.29, No.05, pp.66-70.
[10] Z.Chen and G.C. Luo, 'Improved K- Means Algorithm Using in Anomaly Detection', Chongqing Science University of Technology (NATURAL SCIENCE) **(2015)**, Vol.29, No.05, pp.66-70.
[11] H.Li, 'Implementation of an Anomaly Detection Model Based on Improving Clustering Analysis', Microelectronics and Computer **(2010)**, Vol.27, No.08, pp.66-69.
[12] Z. Muda, W. Yassin, M.N. Sulaiman and N.I. Udzir, 'A K-Means and Naive Bayes Learning Approach for Better Intrusion Detection', Information Technology Journal **(2011)**, Vol.10, NO.3, pp.648.
[13] A. Rodriguez, and A. Laio, 'Clustering by fast search-and-find of density peaks' Science **(2014)**, pp.1492-1496.
[14] J. Wang, J-Uk Kim, Lei Shu, Yu Niu and Sungyoung Lee, A distance-based energy aware routing algorithm for wireless sensor networks, Sensors, 10, 10, (**2000**)
[15] J. Wang, Yue Yin, J. Zhang, S. Lee, and R. Simon Sherratt, Mobility based energy efficient and multi-sink algorithms for consumer home networks, IEEE Transactions on Consumer Electronics, 59, 1, (**2013**)

Authors

Chunyong Yin, he is currently an associate Professor and Dean with the Nanjing University of Information Science & Technology, China. He received his Bachelor (SDUT, China, 1998), Master (GZU, China, 2005), PhD (GZU, 2008) and was Post-doctoral associate (University of New Brunswick, 2010).He has authored or coauthored more than twenty journal and conference papers. His current research interests include privacy preserving and network security.

Sun Zhang Sun Zhang received his bachelor degree in computer science and technology from Nanjing University of Information Science and Technology, China. Now he is studying for his master's degree in there.



Zhichao Yin, he is studying in Nanjing No.1 Middle School. His current research interests include network security and mathematical modeling.

Jin Wang, he received the B.S. and M.S. degree from Nanjing University of Posts and Telecommunications, China in 2002 and 2005, respectively. He received Ph.D. degree from Kyung Hee University Korea in 2010. Now, he is a professor in the Computer and Software Institute, Nanjing University of Information Science and technology. His research interests mainly include routing method and algorithm design, performance evaluation and optimization for wireless ad hoc and sensor networks. He is a member of the IEEE and ACM.