International Journal of Security and Its Applications
Vol. 10, No. 8 (2016), pp.353-364
http://dx.doi.org/10.14257/ijsia.2016.10.8.31

Research on Data Intrusion Detection Technology based on Fuzzy Algorithm

Sheng Zhao, Huishan Han and Xuekui Shi

Xingtai Polytechnic College, Xingtai, Hebei, China, 054000
593826346@qq.com

Abstract

The computer system is becoming more complex and massive network data, which brings great difficulties to the traditional intrusion detection system. Intrusion detection system is an important part of the network and information security architecture, which is mainly used to distinguish the normal activities of the system and the suspicious and intrusion patterns. But the challenge is how to effectively detect network intrusion behavior in order to reduce the false alarm rate and false negative rate. Based on the shortcomings of existing intrusion detection methods, the fuzzy C- means clustering method is proposed to analyze the intrusion detection data, so as to find out the abnormal network behavior patterns. By testing the CUP99 data set, the results show that the IFCA is not only feasible but also accurate and efficient. The improved fuzzy clustering algorithm proposed in this paper can improve the detection rate of intrusion detection and reduce the false detection rate, and can be widely used in intrusion detection system.

Keywords*: IFCA; data intrusion detection; fuzzy algorithm*

1. Introduction

With the rapid development of Internet, the corresponding information security issues have become more and more serious, which has become one of the main constraints of the development of Internet. With the continuous development of hacker attack technology, the current access control, encryption, authentication and other major security measures have been unable to meet the requirements of information security. As more and more government, business, financial institutions and departments will connect their databases to the Internet, online databases are under more and more attack, damage caused by more and more, so the security of network database security is focus. Traditional firewall and data encryption, and other various static protection have been very difficult to meet the requirements of network information security, and intrusion detection system as a kind of positive and active information security defense technology, constitute the network information protection system in a extremely important part. Nowadays, more and more attention has been paid to the research on intrusion detection methods and techniques. Data mining is one of the most advanced and active research directions in the field of database and information decision. Intrusion detection method based on clustering analysis, focusing on the traditional clustering methods, such as distance based clustering method (Portnoy *et. al.*, 1998)[1]. Intrusion Detection Based on unsupervised clustering and support vector machine (Luo Min, 2008)[2]. Supervised anomaly detection method based on clustering (Jiang Shengyi *et. al.*, 2010)[3]. This method is the object to be processed each strictly divided into a class, has the characteristics of either this or that. The same thing belongs and only when ownership of the designated categories. This classification category boundaries are clear and membership value is either 0 or 1, is a hard clustering method, the typical algorithms are k-means algorithm. With in-depth research. It is found that this clustering is more and

ISSN: 1738-9976 IJSIA

more not applicable with fuzzy classification problems. Introduced the concept of fuzzy partition of fuzzy clustering analysis, and the subordinate relationship value from {0,1} two valued logic is extended to [0,1] (Ruspinid, 1969) [4]. Based on the hard clustering K-means algorithm, a partition type fuzzy clustering method based on objective function is proposed. The typical algorithms is fuzzy C-means (FCM).K- mean and FCM two methods are based on clustering algorithm, the algorithm firstly to construct an objective function and through continuous iterative optimization objective function division of the data set is obtained(Dunn and Bezdek, *etc.*, 1974)[5-6].In essence, they are a kind of clustering method based on probabilistic constraints, and the conditions are satisfied with the requirements of $\sum_{i=1}^{c} u_{ik} = 1 \forall k$ (U_{ik} is membership). This kind of method is objective, there exist some problems: one is many practical data set membership and may not satisfy the probability and 1; the second is the FCM algorithm is affected by the impact of noise, usually individual outliers may lead to the clustering center offset. The common fuzzy clustering algorithm is improved, a uncertainty membership degree is the possibility of improved based on fuzzy clustering algorithm and the for anomaly intrusion detection, through the dataset KDDCUP99 experimental verification is presented. The method can obtain higher detection rate and lower false detection rate for anomaly intrusion detection.

2. The Principle of FCM Clustering

Fuzzy C- means (FCM) algorithm is one of the most important and popular fuzzy clustering algorithm.

2.1. Fuzzy C- mean Method

Given data set $X = \{x_1, x_2, \cdots, x_n\} \subset R^s$ is a feature vector set of s dimensional pattern space, according to some similarity measure, the set is aggregated into C sub set $X_1, X_2, \cdots, X_n, 2 \le c \le n$. The C subset is a fuzzy partition of the feature vector set X, μ_{ik} indicates that the feature vector X_i belongs to the membership degree of the subset X_k, thus the fuzzy classification can be got.[7]

$$U = [\mu_{ik}]_{c \times n} = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1n} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{c1} & \mu_{c2} & \cdots & \mu_{cn} \end{bmatrix}_{c \times n} \tag{1}$$

And meet $0 \le \mu_{ik} \le 1(1 \le i \le c, 1 \le k \le n), \sum_{i=1}^{c} \mu_k = 1(1 \le k \le n), 0 < \sum_{k=1}^{n} \mu_{ik} < 1(1 \le i \le c)$. Make a collection of all $U = [\mu_{ik}]_{c \times n}$ records as M_{fcn} . Namely $M_{fcn} = \left\{ U = [\mu_{ik}]_{c \times n} \middle| \mu_{ik} \in [0,1], \forall i, k; \sum_{i=1}^{c} X_{ik} = 1, \forall k; n \ge \sum_{k=1}^{c} X_{ik} \ge 0, \forall i \right\}$, called X M_{fcn} soft C space division. Representing all real $c \times s$ order matrices by matrix R^{cs}. Order $v = (v_1, v_2, \cdots, v_c)^T \in V$ is the cluster center, where $v_i \in V$ is the cluster center of the class $i(1 \le i \le c)$. The objective function of FCM can be expressed as $J_m(u, v)$:

$$J_m(u,v)=\sum_{i=1}^{c}\sum_{k=1}^{n}(\mu_{ik})^m\|x_k-v_i\|^2 \tag{2}$$

Wherein, $\|x_k-v_i\|^2=(x_k-v_i,x_k-v_i),1<m<+\infty$.

Dunn proves that the functional minimum problem is solvable. Bezdek has proved that (u^*,v^*) is a necessary condition for the local minimum of $J_m(u,v)$:

$$v_i^*=\frac{\sum_{i=1}^{n}(\mu_{ik})^m x_k}{\sum_{k=1}^{n}(\mu_{ik})^m}(i=1,2,\cdots,c) \tag{3}$$

$$u_{ik}^*=\frac{1}{\sum_{j=1}^{c}\left(\frac{d_{ik}^*}{d_{jk}^*}\right)^{\frac{1}{m-1}}}(i=1,2,\cdots,c;k=1,2,\cdots,n) \tag{4}$$

At the same time, Bezdek gives the $m\geq 1$ and $x_k\neq v_i$ iterative algorithm. [8]

2.2. Fuzzy C- mean Algorithm

The basic idea of FCM algorithm is to use the iterative method to solve the formule v_i^* and u_{ik}^*, until the termination of a certain conditions, the specific steps are as follows:

Step one: given the number of cluster C, weight index m and iterative standard ε;

Step two:select the initial cluster center $v=(v_1,v_2,\cdots,v_c)^T$;

Step three: using the current clustering center to calculate the membership function u_{ik}^*;

Step four: using the current membership function to update the calculation of all kinds of cluster centers v_i^*;

Step five: if the distance between the fuzzy matrix is not more than ε, then the algorithm terminates, otherwise it will not turn to step 3.

The (U^*,V^*) obtained by the above algorithm is the optimal solution with respect to the number of C, the weighted index m and the ε. Therefore, for different m values, there will be a different fuzzy C- division, it is necessary to consider the problem of the optimal value of M.

2.3. Determination of Optimal Weight Index m^*

The weighting exponent m is one of the FCM algorithm is one of the most important parameters, its value affects not only the the concavity and convexity of the objective function, but also control the clustering of fuzziness, fuzzy between class share degree, noise, the objective function of the concavity and convexity and the algorithm convergence. In theory, M range for $[1,+\infty)$. However, when m is close to FCM, l is degenerate into hard C- mean algorithm. When m tends to infinity, the only solution of FCM is the centroid of the data set, and the partition function is lost. If the M is too small, the anti noise performance of FCM algorithm becomes worse. Conversely if we take too much, not accurate prototyping model. The optimal weight of index should be the fuzzy clustering within class weighted error and the minimum. At the same time, but also to ensure inter cluster good separability. By

$$\frac{\partial J_m(u,v)}{\partial m}=\sum_{i=1}^{c}\sum_{k=1}^{n}\left[(\mu_{ik})^{m-1}\|x_k-v_i\|^2\right]\left[\mu_{ik}\ln(\mu_{ik})\right] \tag{5}$$

It is known that $J_m(u,v)$ decreases monotonically with the increase of M. The research shows that the objective function $J_m(u,v)$ has a very small point on the partial derivative of the parameter m, and it happens to be in the range of $[1,1.5]$. Therefore, an optimal m selection method is obtained:

$$m^*=\left\{m\left|\frac{\partial}{\partial m}\left[\frac{\partial J_m(u,v)}{\partial m}\right]=0\right.\right\} \tag{6}$$

Taking into account the real-time nature of the intrusion detection, the author simplifies the:

$$m^*=\arg_{\forall m}\left\{\min\left\|\left[\frac{\partial J_m(u,v)}{\partial m}\right]\right\|\right\} \tag{7}$$

In this paper, under the condition of selected number c=2, m=2.[9]

2.4. Relative State Characteristic Value

FCM algorithm to get the optimal classification matrix U^* is a fuzzy matrix, the corresponding classification is soft classification. In order to make the U^* clear, the principle of maximum membership degree is generally used, but the principle of maximum membership is sometimes not applicable. If misuse of this principle, it can lead to unreasonable judgment. Therefore, the relative state characteristic value is adopted.[10]

According to the optimal classification matrix U^*, the relative membership degree of the sample x_i to the category $\mu_{1i}^*,\mu_{2i}^*,\cdots,\mu_{ci}^*$ is $l=1,2,\cdots,c$ as the relative membership weight of the state variable L, and the sum is $L(x_i)=\sum_{l=1}^{n}\mu_{li}^*\times l$ as the relative state characteristic value. Visible $L(x_i)$ uses all the relative membership information of the state variable L, and therefore, according to the state of the $L(x_i)$ judgment sample x_i is more precise.[11]

3. Intrusion Detection Technology

3.1. Intrusion Detection System

Intrusion detection system is generally divided into two categories according to the detection method: abnormal intrusion detection system and misuse intrusion detection system.[12]

Abnormal intrusion detection refers to a behavior standard and it is the standard, when the behavior and this standard is not the same time to issue a warning. The key technology of anomaly intrusion detection system is to select the appropriate measure to make the correct judgment between the abnormal behavior and the intrusion behavior. The metrics to be selected depend on the type of intrusion detected, and the different metrics of the intrusion type are also different. A predetermined metric may produce false negatives of intrusion. An ideal anomaly intrusion detection system needs to be able to dynamically make judgments and decisions. Anomaly intrusion detection does not depend on the characteristics of the attack. It is based on the detection of the target to discover the intrusion behavior, it is easy to discover new types of attacks. However, it is difficult to solve the problem of how to establish the normal behavior pattern library and establish the abnormal indexes of the detection behavior. SNORT is one of the most perfect function based misuse detection system. It can not only detect the header information of the network data packets, but also can detect the load data. But it needs to match the network data with all the rules, when the network traffic is very large, it is inevitable that there will be a phenomenon of data packets missing.[12]

3.2. Intrusion Detection Technology Research

Figure 1. Data Intrusion

Intrusion detection can be divided into two types: misuse detection model and anomaly detection model, these two models can be used for real-time detection and ex post detection. The corresponding detection method is misuse intrusion detection and anomaly intrusion detection. The detection process of anomaly intrusion detection is shown in Figure 2, and the detection process of misuse intrusion detection is shown in Figure 3.

Anomaly detection based on normal behavior pattern library. When the difference between the data and the normal value of the system (the user) in real time exceeds a given threshold value, the data can be judged as abnormal data. But in reality, the abnormal behavior is not equal to the intrusion behavior. Generally speaking, the behavior of the following situations: (1) behavior is intrusion behavior, but in the test not exceptional; (2) is the intrusion behavior, performance is also abnormal behavior; (3) behavior is not the invasion behavior, but the performance is abnormal behavior; (4) behavior is not into the intrusion behavior, the performance is not abnormal behavior. Anomaly intrusion detection has the advantages of unknown intrusion data can be found, but due to the complexity of the computer system to to detect user actions and behavior to be all were comprehensive detection, so the detection error rate is relatively high, management is also difficult. Abnormal intrusion detection is needed to select the appropriate feature when establishing the normal profile of the normal behavior. The selection of characteristics should not only reflect the behavior of the system or the user, but also make the model optimization. In the process of intrusion detection, it is the difference between the test data and the normal data to judge whether it is the basis of the intrusion behavior. So the establishment of reference threshold is a key problem. The threshold is set too large, the false negative rate will become large, the threshold is too small, then the error rate will be very high. Therefore, the appropriate threshold setting is a key factor for the accuracy of anomaly detection. In anomaly detection, data mining is the most widely used method. Data mining is a process of finding useful knowledge from a large amount of data. It is a kind of active knowledge discovery technology. Using data mining technology can effectively analyze the data to be detected, extract the characteristics of intrusion behavior, and sum up the corresponding detection model to detect and analyze the same kind of attack behavior. Data mining can not only deal with the massive data, but also can carry on the data association analysis. Therefore, in the aspect of intrusion detection, the detection algorithm based on data mining can play a great advantage. So this article will use this method to carry on the simulation analysis.

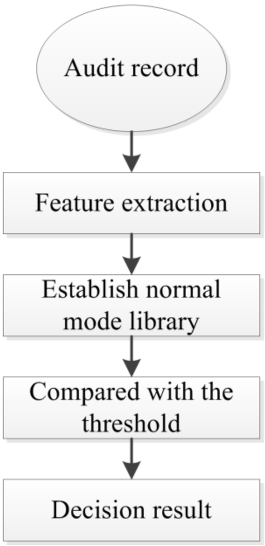


Figure 2. Abnormal Detection Process

Misuse detection is based on knowledge (or feature). It is assumed that all intrusions and means (and variants) can be expressed as a model (or feature), and the method is to detect the intrusion behavior by detecting the matching degree of the known intrusion models. Therefore, we need to understand the attack behavior before the misuse detection. The establishment of intrusion pattern feature database is the key step of misuse detection, which will directly affect the ability of intrusion detection. Misuse intrusion detection has the advantage of less false positives. Disadvantage is that can only be found in the attack in the library of known attacks, the false negative rate high, the system relies on strong, portability is bad and the complexity will increase with the increase of the number of attacks. Because misuse detection does not care about the intent to match behavior, even the normal data sometimes triggers an alarm. Using misuse detection can detect the following types of attacks:

Type one: login failure detection: the number of times the password check failed to detect a specified login activity. Can detect a single user can also be the user of the entire system testing. Many attacks attempt to tend to be in the same account trying to use a password or a password to try different logon account. As a matter of fact, attacks often occur in intensive activities, usually in the form of in a very short period of time, such as a few minutes, a large number of password failure record. When the password failure occurs, the audit data Trace library records: EventClass=20.

Type two: login detection: for the detection of a user in the illegal time or illegal path. The audit records of the Trace library records are: EventClass=14 (Login) and EventClass=15 (Logout)

Type three: the failure detection of the operation record: the number of times the failure of select, Pudat, and delcte, which occur within a given time window (*e.g.,* 5 minutes). When the operation record fails, the system audit trail will record the relevant information. Audit data recorded by the Trace library is: EventClass=33.

Type four: user rights change detection: detection of changes to user rights. When the user rights change occurs, the audit data Traee library records: EventClass=104-112.

Type five: user information detection: for the detection of the user's client program name, user computer name, NT user name, NT domain name and user login information provided when there is access to the police.

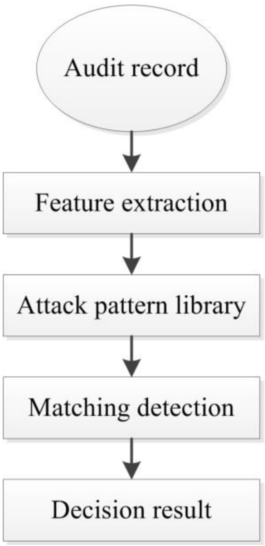


Figure 3. Process of Misuse Detection

The advantage of this method is that it can determine whether the invasion is successful or not, and can determine the success of the previous attacks. Shortcomings can not foresee unknown attacks.

4. Application of Improved Fuzzy Algorithm in Data Intrusion Detection

4.1. Experimental Data

In this paper, we use the standard KDDCUP99 data set, randomly generated from the KDDCUP99 2 experimental data subsets, data set 1.[13]

Consisting of 2048 records, including 2000 normal records, 48 attacks, records of the entire record of the attack rate of 2.3%. The dataset is used as the training set. The data set 2 has 2000 records, including 1889 normal records, 111 attack records. The dataset is used as a test set. In order to verify the detection ability of the anomaly detection method, the data set 2 contains the type of attack that does not appear in the data set 1. The specific data is shown in Table 1.

Table 1. KDDCUP99 Experimental Data

Data set	Types and quantities of attacks

Data set 1	U2R:9(buffer-overflow:6;rootkit:3)
	Probe:12(ipsweep:7;portsweep:5)
	R2L:15(warezclient:11;multihop:2;phf:2)
	DOS:12(smurf:7;teardrop:5)

Data set 2	U2R:6(buffer-overflow:3;rootkit:3)
	Probe:28(ipsweep:4;satan:14;portsweep:10)
	R2L:29(warezclient:3;warezmaster:15;multihop:7;phf:4)
	DOS:48(smurf:9;teardrop:8;neptune:31)

In the original data set, there is an important problem in the form of measurement is different, the direct clustering process will produce not real results, the need for data standardization, that is, data preprocessing. There are 2 steps in the normalization of data: one is the numerical value of the symbolic feature, the KDDCUP99 data set has 3 symbolic attributes. In this experiment, we use the way of distinguishing the protocol layer, such as TCP, UDP, ICMP, and 1,2,3, which appeared in the attribute of the protocol type. In this way, the numerical value of the symbolic attribute is realized, and the other 2 symbolic attributes are also changed. Second is standardization and normalization, to eliminate the dimension effect, the standard deviation and range transform each numeric variable data, make each variable with mean 0 and standard deviation is 1 and each variable in the range [0,1]. The following are given the standard deviation and range of the conversion formula.[14]

$$x''_{ik} = \frac{x'_{ik} - \bar{x}'_k}{s_k} \tag{8}$$

Wherein, $\bar{x}'_k = \frac{1}{n}\sum_{i=1}^{n} x'_{ik}$; $s_k^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x'_{ik} - \bar{x}'_k\right)^2$ 。

$$x_{ik} = \frac{x''_{ik} - x''_{k\min}}{x''_{k\max} - x''_{k\min}} \tag{9}$$

Wherein, $x''_{k\min}$, $x''_{k\max}$ respectively indicates the minimum and maximum values.[15]

4.2. Anomaly Detection

Anomaly detection model based on clustering algorithm, the general requirements of the abnormal behavior of the data in the training set of the proportion of the normal behavior data is 3% ~ 10%. When the cluster set contains the abnormal behavior recorded more than training set the proportion of abnormal records that this collection into a set of abnormal behavior. For the rest of the normal behavior set; anomaly detection is by training a class model, then testing data and categorical model one by one, generally with the detection rate and false detection rate evaluation of the results.

In order to verify the effectiveness of the proposed algorithm, in the PC machine (memory is 256MB, CPU is Pentium2.4GHz, the operating system is WindowsXP) with C++6.0 programming, 3 kinds of clustering algorithms are implemented. All the algorithms have been carried out on the KDDCUP99 data set. First, the data set is trained on the data set, and the results of the clustering are obtained, and then the normal and abnormal behavior pattern library is constructed. This pattern library can be tested on the test data set. The experimental results are shown in Figure 4 and Figure 5.

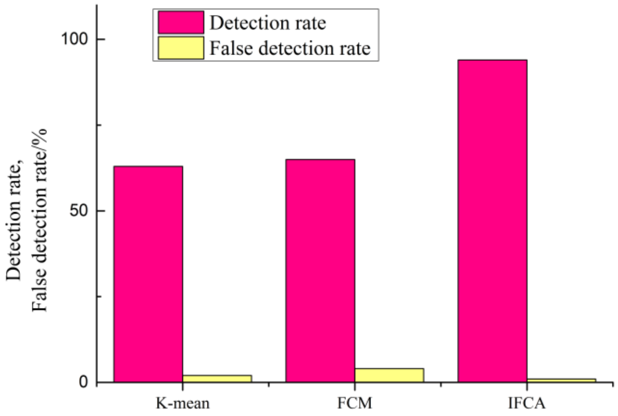


Figure 4.Training Results of 3 Algorithms on Data Set 1

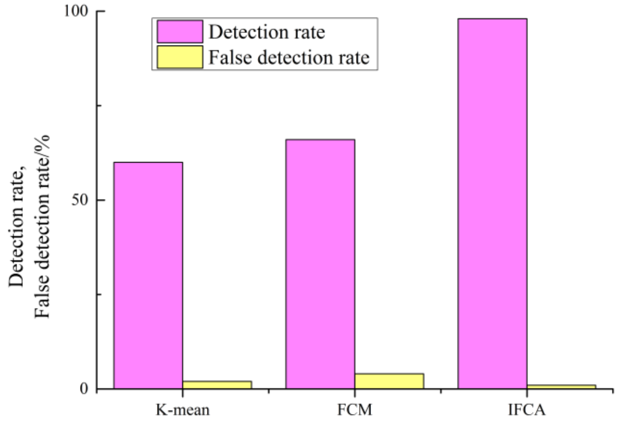
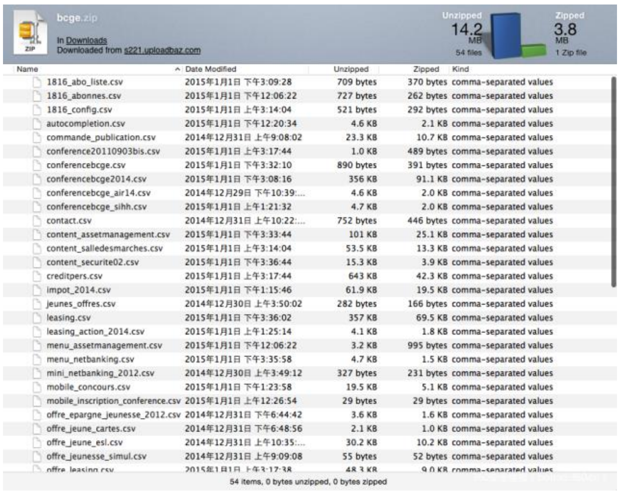


Figure 5. Detection Results of 3 Algorithms on Data Set 2

The experimental results show that IFCA has better clustering effect than the typical K- mean and FCM algorithm. In the intrusion detection, the detection rate of IFCA was 95.8%, higher than the detection rate of K- 60.4% and FCM detection rate of 64.6%. In the case of the same experimental environment and the experimental data set, the results are also higher than Luo Min's experimental method (detection rate is 60%) and Wang Xiaofeng's experimental method (detection rate is 72%-82%). There was no significant difference in the false detection rate. These results show that IFCA can effectively improve the efficiency of intrusion detection, and the effect of anomaly detection is better than that of several clustering methods. The better clustering and detection results are obtained because of the possibility of IFCA clustering and the uncertainty relation. Uncertainty fuzzy clustering algorithm is different from the general fuzzy clustering algorithm, which reflects the uncertainty of the data objects on the cluster membership.

4.3. Misuse Detection

Detection to intrusion detection is mainly: attempted attacks, such as password guessing, unauthorized operation caused by the operation failure, illegal logon time, does not match the user information etc.. Camouflage attack: login through the legal way, but the user's behavior is different from the usual, the operation of the record fails, user rights appear abnormal change (Figure 6). 3 clustering algorithms were used for misuse detection, and the results were shown in Figure 7.



bcge.zip
In Downloads
Downloaded from s221.uploadbaz.com

Unzipped 14.2 MB 54 files
Zipped 3.8 MB 1 Zip file

Name	Date Modified	Unzipped	Zipped	Kind
1816_abo_liste.csv	2015年1月1日 下午3:09:28	709 bytes	370 bytes	comma-separated values
1816_abonnes.csv	2015年1月1日 下午12:06:22	727 bytes	262 bytes	comma-separated values
1816_config.csv	2015年1月1日 上午3:14:04	521 bytes	292 bytes	comma-separated values
autocompletion.csv	2015年1月1日 下午12:20:34	4.6 KB	2.1 KB	comma-separated values
commande_publication.csv	2014年12月31日 上午9:08:02	23.3 KB	10.7 KB	comma-separated values
conference20110903bis.csv	2015年1月1日 上午3:17:44	1.0 KB	489 bytes	comma-separated values
conferencebcge.csv	2015年1月1日 下午3:32:10	890 bytes	391 bytes	comma-separated values
conferencebcge2014.csv	2015年1月1日 下午3:08:16	356 KB	91.1 KB	comma-separated values
conferencebcge_air14.csv	2014年12月29日 下午10:39:...	4.6 KB	2.0 KB	comma-separated values
conferencebcge_sihh.csv	2015年1月1日 上午1:21:32	4.7 KB	2.0 KB	comma-separated values
contact.csv	2014年12月31日 上午10:22:...	752 bytes	446 bytes	comma-separated values
content_assetmanagement.csv	2015年1月1日 下午3:33:44	101 KB	25.1 KB	comma-separated values
content_salledesmarches.csv	2015年1月1日 上午3:14:04	53.5 KB	13.3 KB	comma-separated values
content_securite02.csv	2015年1月1日 下午3:36:44	15.3 KB	3.9 KB	comma-separated values
creditpers.csv	2015年1月1日 上午3:17:44	643 KB	42.3 KB	comma-separated values
impot_2014.csv	2015年1月1日 下午1:15:46	61.9 KB	19.5 KB	comma-separated values
jeunes_offres.csv	2014年12月30日 上午3:50:02	282 bytes	166 bytes	comma-separated values
leasing.csv	2015年1月1日 下午3:36:02	357 KB	69.5 KB	comma-separated values
leasing_action_2014.csv	2015年1月1日 上午1:25:14	4.1 KB	1.8 KB	comma-separated values
menu_assetmanagement.csv	2015年1月1日 下午12:06:22	3.2 KB	995 bytes	comma-separated values
menu_netbanking.csv	2015年1月1日 下午3:35:58	4.7 KB	1.5 KB	comma-separated values
mini_netbanking_2012.csv	2014年12月30日 上午3:49:12	327 bytes	231 bytes	comma-separated values
mobile_concours.csv	2015年1月1日 下午1:23:58	19.5 KB	5.1 KB	comma-separated values
mobile_inscription_conference.csv	2015年1月1日 上午12:26:54	29 bytes	29 bytes	comma-separated values
offre_epargne_jeunesse_2012.csv	2014年12月31日 下午6:44:42	3.6 KB	1.6 KB	comma-separated values
offre_jeune_cartes.csv	2014年12月31日 下午6:48:56	2.1 KB	1.0 KB	comma-separated values
offre_jeune_esl.csv	2014年12月31日 上午10:35:...	30.2 KB	10.2 KB	comma-separated values
offre_jeunesse_simul.csv	2014年12月31日 上午9:09:08	55 bytes	52 bytes	comma-separated values
offre_leasing.csv	2015年1月1日 上午3:17:38	48.3 KB	9.0 KB	comma-separated values

54 items, 0 bytes unzipped, 0 bytes zipped

Figure 6. Interface of Misuse Detection

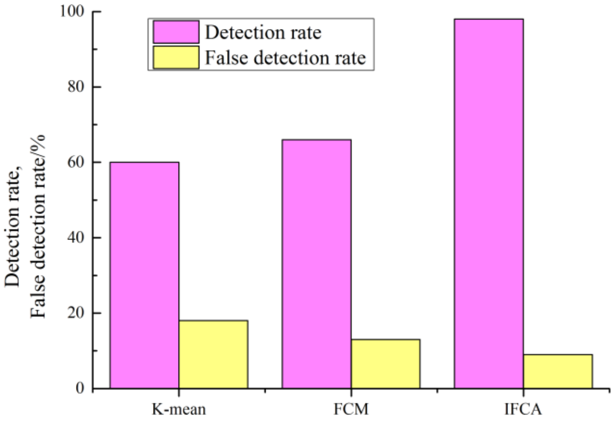


Figure 7.Three Algorithms Misuse Detection Results

From the graph can be drawn, IFCA can effectively improve the efficiency of intrusion detection, the effect of misuse detection is better than the current commonly used in several clustering methods. IFCA false detection rate is relatively low, but still much higher than the abnormal detection. The main reason for this is that the main limitations of misuse intrusion detection is just detection oneself known weaknesses and can only detect the known attack and to detect the unknown intrusion may is of little use, for the intrusion of unknown attack types and internal staff powerless.

5. Conclusions

With the rapid development of Internet, computer network has been playing a more and more important role in the social, economic, cultural and people's daily life. At the same time, more and more government, business, financial institutions and departments will connect their databases to the Internet, online databases are under more and more attack, damage caused by more and more, so the security of network database security has become the focus, we urgently need to research for database intrusion detection technology to improve security. In this paper, a fuzzy clustering algorithm based on fuzzy membership is proposed. In the iterative process, the algorithm creates an uncertain membership degree and a relative membership degree, which makes the elements in the sample not only limited to a cluster. Experimental results on the data set show that IFCA can effectively increase the detection efficiency of intrusion detection system, and the effect of anomaly detection is much better than that of fuzzy clustering algorithm which is often used at present. Because it is the relationship between relative uncertainty and added to the objective function to, so it can get better results, uncertainty, fuzzy algorithm and fuzzy clustering algorithm is not the same. It mainly shows that the uncertainty of data objects and cluster membership. At the same time, the misuse detection is also carried out, and the result shows that IFCA can effectively improve the efficiency of intrusion detection, and the effect of misuse detection is better than that of several clustering methods. IFCA false detection rate is relatively low, but still much higher than the abnormal detection. The improved fuzzy clustering algorithm proposed in this paper can improve the detection rate of intrusion detection and reduce the false detection rate, and can be widely used in intrusion detection system.

References

[1] Sen J., “An Agent-Based Intrusion Detection System for Local Area Networks[J]”, arXiv preprint arXiv:1011.1531, **(2010)**.

[2] Hoque M. S., Mukit M. and Bikas M., “An implementation of intrusion detection system using genetic algorithm[J]”, arXiv preprint arXiv:1204.1336, **(2012)**.

[3] Phua C., Lee V. and Smith K., “A comprehensive survey of data mining-based fraud detection research[J]”, arXiv preprint arXiv:1009.6119, **(2010)**.

[4] Wang G., Hao J. and Ma J., “A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering[J]”, Expert Systems with Applications, vol. 37, no. 9, **(2010)**, pp. 6225-6232.

[5] Wu S. X. and Banzhaf W., “The use of computational intelligence in intrusion detection systems: A review[J]”, Applied Soft Computing, vol. 10, no. 1, **(2010)**, pp. 1-35.

[6] Horng S. J., Su M. Y. and Chen Y. H., “A novel intrusion detection system based on hierarchical clustering and support vector machines[J]”, Expert systems with Applications, vol. 38, no. 1, **(2011)**, pp. 306-313.

[7] Shanmugavadivu R. and Nagarajan N., “Network intrusion detection system using fuzzy logic[J]”, Indian Journal of Computer Science and Engineering (IJCSE), vol. 2, no. 1, **(2011)**, pp. 101-111.

[8] Teng S., Du H. and Wu N., “A cooperative network intrusion detection based on fuzzy SVMs[J]”, Journal of Networks, vol. 5, no. 4, **(2010)**, pp. 475-483.

[9] Jawhar M. M. T. and Mehrotra M., “Design network intrusion detection system using hybrid fuzzy-neural network[J]”, International Journal of Computer Science and Security, vol. 4, no. 3, **(2010)**, pp. 285-294.

[10] Modi C., Patel D. and Borisaniya B., “A survey of intrusion detection techniques in cloud[J]”, Journal of Network and Computer Applications, vol. 36, no. 1, **(2013)**, pp. 42-57.

[11] Gogoi P., Borah B. and Bhattacharyya D. K., “Anomaly detection analysis of intrusion data using supervised & unsupervised approach[J]”, Journal of Convergence Information Technology, vol. 5, no. 1, **(2010)**, pp. 95-110.

[12] Abadeh M. S., Mohamadi H. and Habibi J., “Design and analysis of genetic fuzzy systems for intrusion detection in computer networks[J]”, Expert Systems with Applications, vol. 38, no. 6, **(2011)**, pp. 7067-7075.

[13] Kshirsagar V. K. and Tidke S. M., “Vishnu S. Intrusion Detection System using Genetic Algorithm and Data Mining: An Overview[J]”, International Journal of Computer Science and Informatics ISSN (PRINT), 2231: 5292, **(2012)**.

[14] Ponomarchuk Y. and Seo D. W., "Intrusion detection based on traffic analysis and fuzzy inference system in wireless sensor networks[J]", J Converg, vol. 1, no. 1, **(2010)**.

[15] Davis J. J. and Clark A. J., "Data preprocessing for anomaly based network intrusion detection: A review[J]", Computers & Security, vol. 30, no. 6, **(2011)**, pp. 353-375.

Authors

Sheng Zhao, He is a lecturer of Xingtai Polytechnic College, and he has got the degree of Master.