ATLANTIS PRESS

Advances in Computer Science Research, volume 83

8th International Conference on Social Network, Communication and Education (SNCE 2018)

Network Security Based on K-Means Clustering Algorithm in Data Mining Research

Chunfen Bu[1,a *]

[1]Department of physical science and technology, Kunming University

[a]18459423@qq.com

*The Corresponding Author

Keywords: Intrusion detection systems; Data mining; Network security; Cluster algorithm

Abstract. Nowadays, the network has become the basis of everything. Meanwhile, network security has become one of today's most urgent social problem. Intrusion detection systems are sold through real-time monitoring of network traffic, and take corresponding measures when the suspicious transfer of suspicious problems of a new network security device. Intrusion detection system compared to traditional network security measures, have great advantages. Can solve the shortcomings of the original passive inspired, can also process it before the damage occurred, appearance of the intrusion detection system, has become an important part of network security.

Preface

In today's society, computer network security has become the chief problem of information society. With the continuous development of technology, the network intrusion behavior has the hidden power, the means of destruction is complex, there is no time space to restrict the existence of network, there is a great harm to the network security [1-3]. Therefore, network security is the most important component of today's society. As for the detection and prevention of intrusion detection, it becomes the primary problem that we need to solve. The research on intrusion detection system also becomes extremely important. Based on the data mining of k-means clustering algorithm, this paper conducts research on network security and discusses how to create a network security and harmonious environment [4].

Intrusion detection system is a system that can detect all software and hardware, and the application value is high. At present the system has already become the main network security management tool, can collect different set information in the system, and then combined with the function of the system of detection and automatic response [7]. Intrusion detection system is a behavior classifier, which operates through the judgment of information intrusion and non-invasive behavior. Here is the concept associated with intrusion detection.

In the early intrusion detection system, Denning successfully proposed the general intrusion detection system model [8], which laid a solid foundation for future research of intrusion detection system.

Data Mining Algorithm

Data mining [9-11] algorithm consists of cluster analysis algorithm, correlation analysis and classification algorithm. Clustering algorithms can be the object of the data set is divided into a lot of similar classes, and classification algorithm is similar, are complete data grouping, and then reference algorithm definition, with the help of clustering algorithm can obtain high similarity of the same object.

Cluster analysis is a common method in data mining analysis, which can be used to show unsupervised anomaly detection, and can solve problems existing in traditional data mining methods. This method can be used in a new database without having to rely on pre-determined data categories and data category samples in intrusion detection system. Cluster analysis creates a good environment for the establishment of intrusion detection system.

Intrusion detection system is mainly to distinguish normal behavior and abnormal behavior and then make corresponding measures. In the midst of a data set, can through the simple data preprocessing and system audit, to use these data sets in our system, but this method is only used in simple normal behavior and behavior analysis, premise is to know the difference between the abnormal data and normal data. By clustering algorithm, one group can not distinguish between normal and abnormal data processing, can summarize and find common ground, and then make a distinction. Clustering algorithm. Therefore, the application of unsupervised clustering algorithm in the field of abnormal detection can improve the detection efficiency of intrusion detection system and the practical application value is higher.

In data mining, the main need detailed analysis was carried out on the clustering algorithm, and grasp the methods of use of such algorithm, in the middle of the clustering algorithm, the K - means algorithm is one of the most commonly used and most practical way. Next, we analyze the k-means algorithm.

K - means algorithm first determine the input parameters, the n in the sample data is divided into K class, the same data in a cluster similarity is high, the center of the cluster needs to be from the similarity of data in the group of the lower average.

Establishment of Intrusion Detection Model

Four general intrusion detection model is set up, the first to use collection system, guarantee the connection records in the process of use, and can get clustering analysis of data sets, and then with the help of clustering algorithm distribution connection records, distinguish normal and abnormal connection records. In this study, k-means algorithm was used to complete cluster analysis. Clustering algorithm results in more clustering, so there are some connection records in each cluster. According to the properties of a given connection record, the properties can be used to determine the two kinds of abnormal clustering and normal clustering. The exception clustering represents the clustering of the abnormal connection records, and the normal clustering represents the clustering of the normal connection records.

In system applications, if you can't use tagged data, you can't clearly determine the normal or abnormal condition of the connection record, and then make the clustering tag. Typically, a threshold is used to record the record of the connection above the threshold for the normal clustering, whereas the other is exception clustering. Using cluster analysis result intrusion methods that connection records, first carries on the standardization, and then from the cluster aggregation clustering, to find the right to his central value close to the distance, complete classification operation according to the tag.

Experimental Results

Data mining has a set of complete analytical method [15], mainly adopted detection rate and the rate of false positives as evaluation index, are defined as follows: related to, the higher detection rate, the lower the rate of false positives, it shows that the better the performance of the proposed data mining algorithm to detect. Table 1 is the test result.

Table 1 Data detection results of network security system based on k-means clustering algorithm

The normal data set of the test data set has been misreported as the intrusion detection rate %	The normal data set of the test data set has been misreported as the intrusion detection rate %	The normal data set of the test data set has been misreported as the intrusion detection rate %	The normal data set of the test data set has been misreported as the intrusion detection rate %	The normal data set of the test data set has been misreported as the intrusion detection rate %	The normal data set of the test data set has been misreported as the intrusion detection rate %	The normal data set of the test data set has been misreported as the intrusion detection rate %
false alarm rate	false alarm rate	false alarm rate	false alarm rate	false alarm rate	false alarm rate	false alarm rate
Data set 1	1200	25	22	9	88	0.75
Data set 2	1220	46	41	10	89.13	0.82
Data set 3	1210	32	27	9	84.34	0.74
Four data sets	1240	64	57	9	90.48	0.73
Data set 5	1230	57	52	10	91.23	0.81
On average,	1220	44.6	39.8	9.4	89.24	0.77

Analysis of table 1, in a given data set, we average detection rate of 89.24%, average 0.77%, the rate of false positives and the detection rate and the rate of false positives there are no big fluctuations, the correctness of the algorithm is applied in the network security system is verified.

Conclusion

The vast amount of data generated in the Internet era undoubtedly challenges the technology of large-scale data processing and data mining. In this paper, we study the problem of network security by using k-means clustering algorithm in data mining. Analyses the network security problems and performance better intrusion detection system in network security analysis simulation, let more people know the network intrusion behavior produces a variety of ways and means. In this way, we can ensure the security of the network information in the network information leak serious today.

References

[1] Hartigan,JA.Clustering Algorithm[M].New York:John wiley &sons Inc.1975.
[2] McQueen show J.S ome the Methods for Classification and Analysis of Multivariate Observations [C]. In: Proceedings of the Sth Berkeley Symposium on Mathematical Statistics and aim-listed Probability. Berkeley, University of California Press, 1967:281-297.
[3] Zheng Yanjun. Application of data mining technology in network security [J]. Computer simulation, 2011(12).
[4] Wang fenglei. Application of improved clustering algorithm in intrusion detection [J]. Computer knowledge and technology, 2011(27).

[5] fuhai Chen. Application of cluster analysis algorithm based on data mining technology in abnormal intrusion detection [J]. Software guide, 2011(5).
[6] Jiang Shengyi, li qinghua, wang hui et al. A guided intrusion detection method based on clustering [J]. Miniature microcomputer system, 2005,26(6):1042-1045
[7] Yan Xinjuan, tan minsheng, yan ya zhou. Research on intrusion detection based on the hidden markov model and neural network [J]. Computer application and software, 2012, (2).
[8] Xiao Jun. The detection method of telecom network intrusion based on the hidden markov model [J]. China science and technology expo, 2010,(12).
[9] Zhang Songhong, wang ydi, han jihong. Research on the prediction of compound attack based on attack intention [J]. Computer engineering and design, 2007, (21).
[10] Creator R, Kaufman L,Source R,et a1.K-means Type Algorithms:A Generalized Convergence Theorem and Characterization of Local Optimality[J].IEEE Transactions on PatternAnalysis and Machine Intelligence,1984.6(1): 81, 87.
[11] Xu L,Krzyzak A,Oja E.Rival penalized competitive learning for clustering analysis,RBF net, and curve detection[J]. Neural Networks,IEEE Transactions on,1993,4(4):636-649.
[12] Zhang dajun, li yunfa, zheng zhou. Security sharing mechanism of data resources in cloud computing [J]. Information network security, 2012,(08):79-82.
[13] liu Xuefei, wang xuefei, wang shenqiang. Implementation of network line data traffic monitoring [J]. Information network security, 2012, (11) :60-62.
[14] Huang Jianwen, tian hongqiang, pei jian. Exploration and practice of data security protection system for operator user data [J]. Information network security, 2012, (12) : 80-82.
[15] Wang Xizhong, qu jiaxing, huang junqiang et al. Implementation of network database security detection and management programming [J]. Information network security, 2012, (02) : 14-18.