Research on Data Stream Technology in Computer Network Security Monitoring

XIONG Wei[1, a]

[1]Chongqing College of Electronic Engineering, Chongqing 401331,China

[a]xiongwei@163.com

Keywords: Data Stream Technology; Computer Network; Security Monitoring

Abstract. With the rapid and bursting development in the subject of computer science and data mining technology, the combination of data analysis and network security is indeed needed. In response of the fact that traditional intrusion detection systems are not able to fulfill the requirements for specific network security, such as fast processing speed, stronger defense capability, and higher real-time performance. We propose a model based on network security on the integration of data stream mining and intrusion detection, after comparing data stream clustering methodology with other state-of-the-art clustering algorithms, we choose the former to act as the selected clustering algorithm. Finally, with the case analysis and simulation experiment, we illustrate the robustness and effectiveness of our proposed methodology.

Introduction

Intrusion detection technology as a supporting mechanism for network security can monitor and detect unauthorized network usages or abnormal conditions without affecting network performance, and then counter such phenomena, thereby attaining the aim of network security defense [1]. With the improvement of network performance as well as diversification of attack technology, however, the network security defense of traditional intrusion detection systems is faced with new challenges. In 1999, from the computer intrusion detection system research of Li science department, Columbia University, the application of data mining technology applied to intrusion detection systems for the first time. This research project is a part of the DARPA for the U.S. Department of defense. Experiments show that this method can improve the detection rate without prejudice to any other model performance test system. In the project of MADAMID (system data mining intrusion detection framework construction characteristics and model of the Columbia University in the United States), a kind of method, more automatic than by manual information system engineering, through construction of an intrusion detection system and data mining technology. Association rules and frequent episode rule proposed remedies more and more predictable characteristics.

In China, Graduate School of Chinese Academy adopted the hierarchical cooperation as a model in an effort to analyze security audit data with data mining algorithms and help the system not only automatically generate intrusion detection rules but also establish anomaly detection models. Tsinghua University came up with a framework for collaborative intrusion detection system (COIDS) based on the method of data mining, and adopted a three-layer (Agent/Manager/UI) entity structure to establish a detection model by various methods of data mining. In fact, Data Mining Intrusion Detection Based on the many achievements have been made, some of which even in use to a certain extent. In such a feature is the larger amount of data for network transmission, faster speed, more means of access, network protocol update ceaselessly, how to improve the protocol intrusion detection system of recognition and processing will be a new problem, it has become a hot topic research in network security question. In this paper, we present a novel network security model built on the integration of data stream mining and intrusion detection system [2].

Data Stream and Network Security Monitoring

Data Stream Mining. Data streams continuously flow into and out of network systems at different rates. These data are characterized as being chronological in order, rapid in change, enormous in capacity, and unlimited in potential. Data stream mining is possessed of the characteristics as follows: With an infinite growth in amounts of data, data stream is limited to single-pass scanning, i.e. every data stream is processed only once unless it is deliberately saved. The need for rapid flow of data flow speed, algorithm analysis of data flow is not lower than the flow of data processing, therefore, the data stream algorithm time is strictly limited. With an unlimited amount of data flow, system memory problems, in this light, there is a complex space requires a higher degree of data stream mining algorithm. The amount of data in the data stream is infinite, so that the storage of data stream mining is impossible data. In order to make up for it, what is to keep the summary information memory of original data, the final results are generated based on. Therefore, data stream mining is often the result of approximation. In the dynamic environment of data stream, the distribution model of underlying data will be affected by a variety of factors from real-world applications. And with the continuity of time, data streams usually keep evolving. Thus, there are usually many conceptual models in data stream mining.

Network Security Monitoring. Intrusion detection system (IDS for short) is a system that detects a variety of intrusions. As an important part of the network security system, it serves to ensure confidentiality, integrity and availability for system resources through. monitoring the running state of network systems and identifying attack attempts, attack behavior and attack results in time to respond accordingly. The system integrates information the interests of their own, silent, passive, in the inlet end through the network, only one listener ports do not forward any traffic. Information collection based on intrusion detection system to extract the statistical characteristics of the corresponding value of the first data stream, and then analysis meets the intrusion knowledge base of these features with the built-in. According to the preset threshold, the packet flow, in the matching and coupling of high, be regarded as aggressive behavior; similarly, the intrusion detection system will come to set the alarm or start a limited counterattack. In the following figure.1 we show the basic steps:

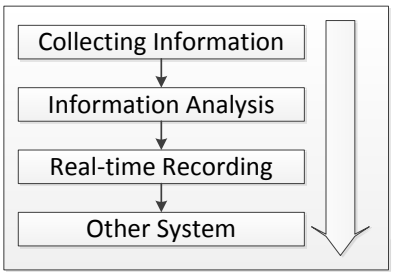

Fig. 1 The Workflow of Network Security Monitoring System

Our Proposed Model and Analysis

When flowing into the network, external information is collected by the collection device first, and then the collected data go through data mining with the data mining algorithm. According to certain rules, the input characteristics of information are aggregation (or classification) and invasion in the existing information systems. If the match is successful, the input information must contain the information of intrusion. In this case, the system alarm, prompting time administrator to handle it. If the match is not successful, this indicates that, the input information is normal behavior, without special treatment. Of course, the matching rules of regularly updated system. As described in Figure 2, the intrusion detection model is based on data stream mining[3].

Data Stream Collection Module. Data collection module is mainly responsible for the lossless capture of network packets, and meanwhile in charge of some simple packet inspection (such as IP packet version checking, validation, protocol inspection, etc.) as well as filtration of error messages. The data the data collection module submits to the pretreatment layer are basically the original data packets. In the previous data collection, the data streams in the network are normal data streams for some time, which, though with no signs of attacks, may have interference noise. As a set of training data, these data provide data for the normal model mining.

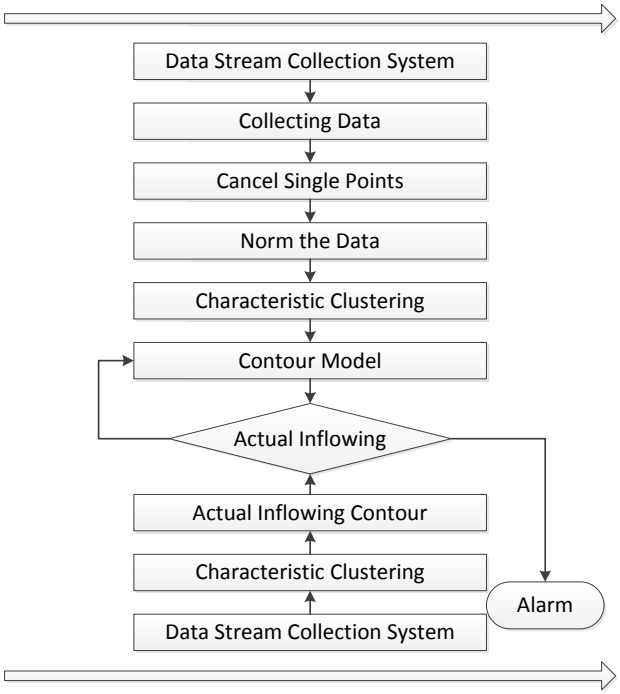

Fig. 2 The Real-time Network Data Stream

Data Stream Collation Module. This module includes training data collation and isolated point exclusion. It is mainly used in the course of data collection, during which the information contains some common processing operations, but with intrusion information excluded as a premise. The data collection mainly refers to the lossless capture of messages. Whether or not the collected data can serve as desired training data depends on the quality of data during the period of collection.

Characteristics Clustering Algorithm Module. As an important technique of data mining, cluster analysis enjoys a broad application. The data set object is divided into several groups of cluster analyze the data in each group were as high similarity, making the group between the lowest possible similarity. Data mining algorithm for extracting the attribute characteristics of safety related system, and then generate a classification according to the security event model these properties, so as to effectively reduce the uncertainty caused by human factors the intrusion pattern analysis feature extraction, thus achieving an automated screening of security incidents.

Intrusion Detection Module. The intrusion detection model based on data stream mining contains such functional modules as data collection, pretreatment, characteristic variable selection, algorithm comparison, mining results treatment and results visualization, among which the data-mining-based algorithm is the core of the whole modeling process.

Implementation and Experiment Analysis

As mentioned above, the construction of detection model for data mining is the core of the whole modeling process [5]. When a new data point p arrives, it is merged, the process of which is described as follows from step one to step three.

Step One. An attempt is made to merge the newly arrived data point P into the candidate micro-cluster C_P closest to it. If the new radius r_P after the merger is less than or equal to the specified threshold ε_1, this point P can be really incorporated into it. The merging operation can be effected based on the incremental maintenance nature of candidate micro-clusters.

Step Two. If the step above fails, an attempt is made to merge the point P into the critical candidate micro-cluster c_0 closest to it. If the new radius r_0 of c_0 after the merger is less than or equal to the corresponding threshold ε_2. If its weight w is greater than $\beta\mu$ that means c_0 has grown into a micro-cluster. In which case, c_0 is removed from the critical candidate cache; and a new critical candidate micro-cluster is established in terms of c_0.

Step Three. If the two steps above are still unable to merge data point P a new critical candidate micro-cluster c_0 is established in terms of P and c_0 is then inserted into the critical candidate micro-cluster cache, waiting for the follow-up treatment. The following table1.1 shows the result.

Data Types	Sample Number	Percentage1 (%)	Test Samples	Percentage2 (%)
Normal	4743	67.76	1867	62.23
Dos	786	11.23	345	11.5
R2L	653	9.32	289	9.63
U2R	456	6.51	168	5.60
Porbing	362	5.17	331	11.03
Total	7000	100	3000	100

Table. 1 The Experimental Result

Summary and Conclusion

As for the conclusion, the objectives for this project were achieved and functioned well as the desired target. Intrusion detection technology as a supporting mechanism for network security can monitor and detect unauthorized network usages or abnormal conditions without affecting network performance, and then counter such phenomena, thereby attaining the aim of network security defense. In this paper, we analyze the basic theory of network security and related subjects, conducting experimental and theoretical approach to verify the proposed model. The result show the robustness and effectiveness of the propose method. In the future, we plan to use more mathematical analysis to design more models.

References

[1] Xu Yonghong, Yang Yun, Cao Lixin, et al. Design of Data-Mining- Based Intrusion Detection System [J].Computer Engineering and Application.

[2] Ji Lei. Research into Data-Mining-Based Intrusion Detection Technology [D].Shanghai, Shanghai Jiaotong University, 2007.

[3] Tang Zhengjun. Design and Implementation of Network Intrusion Detection System [J].Computer Application Research.

[4] Haoxiang Wang. "CLASSIFYING GRAY-SCALE SAR IMAGES: A DEEP LEARNING APPROACH", Machine Learning and Applications: An International Journal (MLAIJ) Vol.1, No.1, September 2014.

[5] Lee W, Stolfo S J, Mok K WA. Data Mining Framework for Building Intrusion Detection Models [C].IEEE Symposium on Security and Privacy. 1999, 13(1):120-132.