2010 Second International Workshop on Education Technology and Computer Science

Research on Anomaly Detection of Network Traffic Based on Fractal Technology and Vector Quantization

Mingsheng LIU[1,2] , Yuemei HE[1,2] , Qingli MENG[2] , Zhihui WANG[1,2]
1 Department of Computer Science, Handan College, Handan 056005, China

2 School of Information & Electronic Engineering, Hebei University of Engineering, Handan 056038, China
hymhdxy@163.com

***Abstract*—In this paper, with the research on the development survey of the network anomaly detection at home and abroad, a new algorithm for anomaly detection of network traffic based on fractal technology and vector quantization is proposed in view of most anomaly detection model with the poor real-time, the lower detection rate and the higher false positive rate. Theoretical analysis shows that this algorithm can achieve higher precision with less space and time complexity, and it can accurately and effectively discover the abnormal network traffic and identify the cause of anomaly network traffic.**

***Keywords*—fractal; vector quantization; network traffic; anomaly detection**

I. INTRODUCTION

With the rapid development of information technology, network-based applications and services have been used more widely. But the network intrusions and attacks have been become more pervasive and devastating, which threaten the security of Internet. It is so imperative that the robustness and security of network should been enhanced for network services and applications.

It has been for ten years since Denning presented anomaly detection model as a branch of intrusion detection. A large number of algorithms and models of computer science areas have been used in anomaly detection of network traffic, including machine learning, data mining, clustering analysis, support vector machine, time series and wavelet analysis. However, with the increase of network speed and network data that need to process, there are still some problems for some detection models: (1) Although traditional machine learning and data mining algorithms have been applied in anomaly detection widely, these algorithms and models are mostly just a simple application and can not meet the requirements of real-time detection algorithm[1]. The majority of algorithms have poor real-time, particularly for the malicious flow and the high-bandwidth data streams. (2) It is very difficult to balance the relationship between the detection rate and the false positive rate for most anomaly detection model with the poor real-time, the lower detection rate and the higher false positive rate, which they can't reduce the false positive rate at the same time they achieve the higher detection rate.

In recent years, research on network anomaly detection based on fractal technology and vector quantization are two hot topics for network anomaly detection. In 2005, Bulut proposed an efficient mutation detection algorithms[2], the algorithm indexed the aggregate calculations results of a few of sliding windows by maintaining minimum layered and limited rectangles. However, the mutation detection method [2,3] can only discover the muation of the monotonic aggregation function of sliding window but can't find the mutation of the non-monotonic aggregation function. References [4,5] are also proposed the detection methods for the mutation behavior of network traffic. These different methods have the same problem commonly, which they can only detect sudden changes at the same time for a single window. In fact, it is very difficult to determine the length of the sliding window for the practical applications. Therefore, in order to find all the mutations and meet all the requirements of network anomaly detection, the best simple way is to perform some algorithms separately, which will obviously lead to the higher space and time complexity and the lower of system performance.

Reference [6] proposed a mutation detection method over data streams based on fractal technology. The algorithm can achieve higher precision with less space and time complexity as compared with the existing methods, and it could be concluded that the proposed algorithm is suitable for burst detection over data streams.

Reference [7] proposed an anomaly detection method based on vector quantization for the malicious of network. In this method usage profile of network traffic behavior has been constructed by using the idea of data compression and vector quantization. Theoretical analysis and experimental results show that this algorithm is special effective.

II. ANOMALY DETECTION OF NETWORK TRAFFIC BASED ON FRACTAL TECHNOLOGY

Generally the self-similar phenomenon existing in nature is called as fractal. And fractal model is applied to describe the complex shapes of nature [8]. A large number of studies show that the self-similar features exist in nature commonly. For example, the communication behavior on the network, stock data, the physiological phenomena of human and meteorological data are all self-similar. Equation (1) show the power-law scaling relations, it is the important properties of fractal model and is a powerful mathematical tool for describing self-similar feature. In (1), s is the linear

$$q = p \cdot s^{d} \qquad (1)$$

$$\log q = \log p + d \cdot \log s \qquad (2)$$

Supported by the Education Research Project of Hebei under Grant No. Z2009408

978-0-7695-3987-4/10 $26.00 © 2010 IEEE
DOI 10.1109/ETCS.2010.544

characteristic degree, such as time and length; q is the measure based on s, such as the aggregate calculation; p is the proportional constant; d is the flexibility exponent. We can get Equation (2) when both ends of (1) are been converted into logarithmic forms.

For the accurate fractal of data, the value of q always complies with (1) for any length s. And the power law scaling relationship is the necessary and sufficient condition for the determining whether one thing is accurate fractal. Therefore, for the accurate fractal data, all points (logs, logq) in the straight line that is been indicated by (2), d is the slope rate of the straight line. As for the approximate fractal data, the value at both ends of (2) only have the same distribution, d is the slope rate of the regression line for all points (logs, logq).

Piecewise fractal model is more suitable to describe the data of approximate fractal characteristics in the practical application，and has been successfully applied to compress data and construct model for the given data set. Piecewise fractal model are made up of some contraction mappings (M_i: i = 1… k). Each of M_i is corresponds to P_i and P_i', which is the form such Equation(3). It's mathematical foundation is recurrent iterated function system.

$$M_i\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a_{11}^i & a_{12}^i \\ a_{21}^i & a_{22}^i \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} b_1^i \\ b_2^i \end{bmatrix},\ i = 1,2\cdots,m \qquad （3）$$

In the (3), a_{22}^{i} is called the shrinkage factor and $\left|a_{22}^i\right| < 1$. To enable the mapping is single-valued mapping, set $a_{12}^i = 0$.

Based on the theory of fractal, Shouke Qin in Fudan University firstly proposed the fractal-based algorithm for burst detection over data streams (FB algorithm) in his dissertation for doctoral degree. In view of the real-time, infinite and continuous characteristics of data streams [9], FB algorithm in constructing model and detecting anomaly data based on fractal technology mainly includes the following:

Firstly, the monotonic search space building algorithms are proposed for detecting the mutation of monotonic aggregation function and non-monotonic aggregation function for multi-window. The monotonic search space is constructed by sorting all the disorderly sliding windows. If we have detected the mutation of aggregation function on a sliding window, we will conclude that there are the mutations of aggregation function on the sliding windows that have been arranged before it. Therefore we can detect mutations in the monotonic search space by using the binary search algorithm, and can achieve the lower time complexity (O(logm), m is the number of window being monitored).

Then, the piecewise fractal model is presented for the real life data, which consumes O(nlogn) time and O(logn) space to model a n length stream. Fig.1 shows two contraction mappings (M_i and M_j). In order to optimize the algorithm, the error can be limited to a smaller range by parameters.

At last, based on this model a burst detection algorithm is proposed which can provide accurate burst detection for several aggregate functions at any length of sliding window.

Theoretical analysis and experimental results show that the algorithm can achieve higher precision with less space and time complexity as compared with the existing methods，and it could be concluded that the proposed algorithm is suitable for burst detection over data streams.

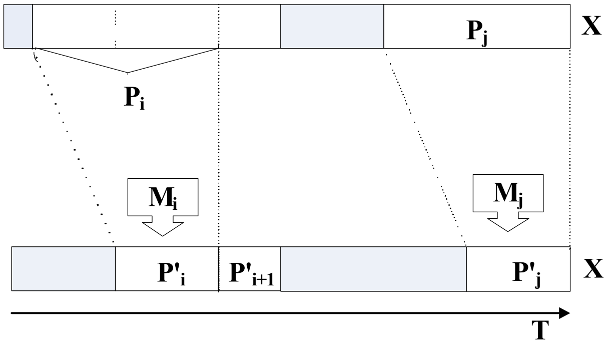

Figure 1. Contraction Mapping of Piecewise Fractal Model over X Data Streams

III. NETWORK ANOMALY DETECTION BASED ON VECTOR QUANTIZATION

Network anomaly detection is an important way to detect the malicious flows, but there are some difficulties in constructing the usage profile of network traffic. The usage profile of network can not been constructed accurately so that most anomaly detection models have the lower detection rate and the higher false positive rate.

Vector Quantization is an important technology of signal encoding and data compression[10]. It has usually been used to compress images and transmit images, as well as search the high-dimensional multimedia files, etc.[11] The input sample space are divided into some signal vectors by the similarity of the input signal data, and each of them replace a vector quantization unit.[12]

Jun Zheng, who studied in institute of technology, Harbin University, proposed the network anomaly detection based on the vector quantization(VQ method) in his dissertation for doctoral degree.

VQ method is an effective anomaly detection method. The construction of network traffic usage profile is first important in anomaly detection. In this paper the data compression and VQ method are proposed to analysis the usage patterns of network traffic. By the similarity partition of network traffic usage space, the codebook of vector quantization can describe the network traffic usage profile accurately and achieved the usage space index via the VQ. The space transfer from network traffic to usage profile is shown in Fig.2.

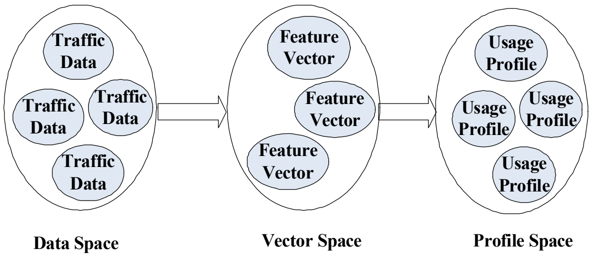

Figure 2. Space Transfer

Because of the amount of data that need to process in network is quite huge, the network anomaly detection needs more efficient algorithms. The fast Nearest-Neighbor search algorithm was proposed to accelerate the process of anomaly detection. It can reduce the unnecessary vector similarity calculation, which can save a great amount of computing time, and achieve the real-time performance of anomaly detection system. It is very important for network anomaly detection.

At last, the efficiency of the method is verified through the evaluating and analyzing the system's performance.

IV. ANOMALY DETECTION OF NETWORK TRAFFIC BASED ON FRACTAL TECHNOLOGY AND VECTOR QUANTIZATION

The network traffic usage profile is constructed by the similarity partition of network traffic usage space, so VQ method can describe the network traffic usage profile accurately and identify the reason of network anomaly with the higher precision and lower false positive rate. However, with the rapid increase of network speed and network traffic, this method should be guaranteed that it can analyze the finer-grained network data, which will inevitably lead to process and analyze more network data even will deduce the efficiency of anomaly detection of network traffic.

FB algorithm can achieve higher precision with less space and time complexity and is suitable for burst detection over data streams, for example, trend analysis, financial risk analysis, communication network monitoring, network traffic management, web log analysis, network intrusion detection and sensor network management, etc. However, the method hasn't further study in identifying the cause of network anomaly.

In order to achieve higher precision with less space and time complexity and to discover the abnormal network traffic and identify the cause of anomaly network traffic accurately and efficiently. This paper presents a new network traffic anomaly detection method: anomaly detection of network traffic based on fractal technology and vector quantization. It is shorted for the FB-VQ method.

With this method we firstly preprocessed the network data in coarse-grained by FB algorithm for which have higher precision with less space and time complexity so that we can remain more useful features for more accurate network anomaly detection in next step. Then we detect the abnormal network traffic by using VQ method. The corresponding code word for each type of attacks are built with more detailed characteristics of data flow in order to we can locate the reason of network anomaly accurately. As shown in Fig.3 :

It is a challenging issue for analyzing the mass network traffic data. By using the FB-VQ algorithm we just with the more finer-grained method analyze the features of the anomaly network data that those have been found with the coarse-grained method, so we can conclude the position and reason of anomaly network traffic accurately and improve the precision of network anomaly detection.

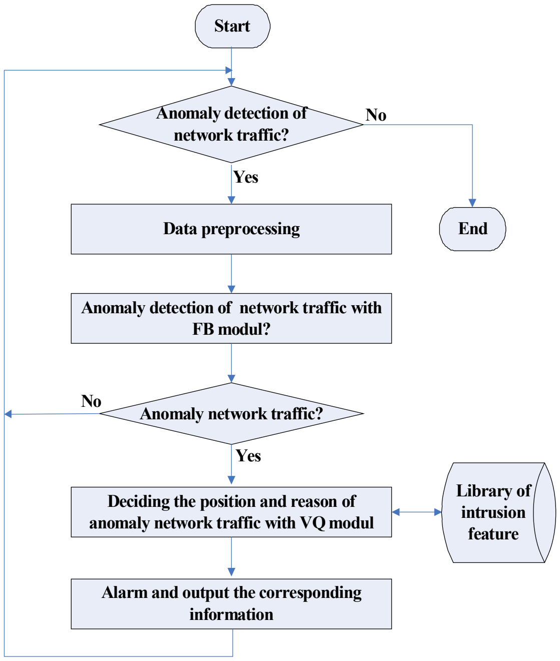
Start
Anomaly detection of network traffic?
No
Yes
Data preprocessing
End
Anomaly detection of network traffic with FB modul?
No
Anomaly network traffic?
Yes
Deciding the position and reason of anomaly network traffic with VQ modul
Library of intrusion feature
Alarm and output the corresponding information

Figure 3. FB-VQ Model of Network Anomaly Detection

V. SUMMARY

In this paper, we firstly study the fractal–based algorithm for burst detection over data streams and the network anomaly detection method based on vector quantization, and then a new network traffic anomaly detection method is proposed. This method bases on fractal technology and vector quantization, it is especially suit for the detecting the malicious of network. It can achieve higher precision with less space and time complexity, and can accurately and effectively discover the abnormal network traffic and identify the cause of anomaly network traffic. Theoretical analysis shows that the method has certain innovation and can been served as a reference for the research of network anomaly detection.

REFERENCES

[1] Juan M. Estévez-Tapiador, Pedro Garcia-Teodoro, Jesus E. Diaz-Verdejo,Anomaly Detection Methods in Wired Networks: a Survey and Taxonomy.Computer Communications. 2004, 27(16): 1569 — 1584

[2] Bulut A,Singh AK.A unified framework for monitoring data streams in real time.In:Kawada S,ed.Proc. of the 21st Int'l Conf. on Data Engineering(ICDE 2005).Tokyo:IEEE Computer Society,2005,44-55

[3] Zhu YY,Shasha D. Efficient elastic burst detection in data streams.In:Getoor L,Senator TE,Domingos P,Faloutsos C,eds.Proc. of the 9th ACM SIGKDD Int'l Conf.on Knowledge Discovery and Data Mining.New York:ACM Press,2003,336-345

[4] Cormode G,Muthukrishnan S. What's new:Finding significant differences in network data streams.IEEE/ACM Trans.on Networking,2005,13(6):1219-1232

[5] Krishnamurthy B,Sen S,Zhang Y,Chen Y.Sketch-Based change detection:Methods,evaluation,and applications.In:Crovella M,ed.Proc.of the 3rd ACM SIGCOMM Conf.on Internet Measurement.New York:ACM Press,2003,234-247

[6] Shouke Qin, Weining Qian, Aoying Zhou. Fractal-Based Algorithms for Burst Detection over Data Streams. Journal of Software, Vol. 17, September 2006, pp. 1969-1979 , in Chinese

[7] Jun Zheng, Mingzeng Hu. An Anomaly Intrusion Detection System Based on Vector Quantization. IEICE Transactions on Information and Systems, Vo1.E89-D, No.1, 2006,201210 (SCI Index 007ET, EI Index 06079701238)

[8] Mandlebrot BB.The Fractal Geometry of Nature.New York:Freeman,1982.37-47

[9] Jin CQ,Qian WN,Zhou AY.Analysis and management of streaming data:A survey.Journal of Software,2004,15(8):1172-11, in Chinese

[10] A. Gersho, R. M. Gray. Vector Quantization and Signal Compression. Kluwer Academic Publishers, Boston, 1992.

[11] Jamshid Shanbehzadeh, Image Retrieval Basedon Index Compressed Vector Quantization. Pattern Reorganization. 2003, 36(11): 2635-2647

[12] K. Goh, E. Chang. Indexing Multimedia Data in High-dimensional and Weighted Feature Spaces. Proceedings of the 6th Visual Database Conference, Australia, 2002: 345 一 356