Large-scale Communication Network Behavior Analysis and Feature Extraction Using Multiple Motif Pattern Association Rule Mining

Weisong He, Guangmin Hu, Xingmiao Yao
Key Lab of Broadband Optical Fiber Transmission and Communication Networks
University of Electronic Science and Technology of China
No.4, Section 2, North Jianshe Road, Chengdu,
CHINA
heweisong@gmail.com

Abstract: - Minimize false positive and false negative is one of the difficult problems of network behavior analysis. This paper propose a large-scale communications network behavior feature analysis method using multiple motif pattern association rule mining, analyze multiple behavior feature time series as a whole, produce valid association rules of abnormal network behavior feature, characterize the entire communication network security situation accurately. Experiment with Abilene network data verifies this method.

Key-Words: - Network behavior analysis, principal components analysis, time frequency analysis, multiple motif pattern association rule mining.

1 Introduction

Network behavior anomalies has the feature of erupting suddenly without known signs, which can bring great damage to network or computers in network in a short time. Therefore, one of the prepositions to ensure a safe network is to detect network behavior anomalies fast and accurately, determine the reasons that causes them and make reasonable response to them in time. To decrease abnormal network behavior and reduce or eradicate attack of denial of service, large-scale communication network routing and switching equipment must possess abilities of detecting and analyzing network behavior behavior. The features of large-scale communication network behavior are high speed and immense data while anomalous behavior is small and scattered in multiple links, which is hard to be detected among normal behavior. Moreover, parameters for analysis are limited. All these make anomaly identification very difficult.

Large-scale communication network behavior anomaly detection is mainly on three levels: packet level, flow level and network-wide level. The advantage of packet level behavior analysis is to provide elaborate information about user performance on the finest level of granularity and basic information about application layer, which in favor of description of anomalous features and fault diagnosis. For example, Snort[2], a signature-based intrusion detection system, summarize packet content in which special attack will appear as one attack feature in artificial ways, and this special attack can be determined if a packet has the same feature when intrusion detection system matches packet content. Ke Wang and Salvatore J.Stolfo [3] adopt the way of using statistical distribution of bytes ASCII code of packet to distinguish the content difference between normal packet and abnormal packet, then the normal connects and abnormal events. Masaki Ishiguro [4] distinguishes network worms attack or port scan through observing packet frequency to specific IP address by Bayesian classification method. But, because of the features of wide distribution, high speed and massive data possessed by large-scale communication network, capturing packet is hard to be implemented on large-scale network. Flow level behavior analysis is based on flow classification, collecting statistical information of each flow and providing performance information of users on medium granularity, which makes characterizing, detecting, diagnosing and restoring network convenient. The idea of flow level behavior analysis is to separate events, group abnormal types, search distributive model of anomaly and analyze anomaly pattern. Since *Netflow* is a good compromise for behavior analysis based on SNMP and packet in its performance and accuracy, the mainstream method of flow level behavior analysis is to be based on *Netflow*. Network-wide behavior analysis uses the global behavior information, including path behavior, link behavior information, etc. for

example, [5] deploys method of multi-way subspace to identify links with behavior anomaly.

Early network behavior analysis mainly focused on the laws that single behavior feature (such as value of behavior, counts of bytes) changes. But single behavior feature time series can not characterize large-scale communication network behavior completely and accurately, so problems like high false positives and false negatives can not be avoided. Lakhina [5] adopted method of multi-way subspace to detect and identify links with behavior feature anomaly. However, this study only used clustering analysis to obtain types of anomaly and did not study correlations among multiple behavior features.

Our contributions lie in: (1) aiming at behavior features of large-scale communication network, make every behavior feature simple time series; then take multiple behavior feature as a whole to analyze and study through multiple motif pattern association rule mining. (2) search motif correlation pattern among anomalous segments of multiple time series within the same time interval by Multiple Motif Pattern Association Rule Mining(MPARM for short in the following), analyze correlation patterns of multiple behavior feature anomaly and describe network security situation of large-scale network accurately and qualitatively.

The rest of the paper is organized as the following. In Section II, we illustrate the process of network behavior analysis and three-level network behavior analysis frame. In section III we illustrate data preprocessing. In section IV we introduce multiple motif pattern association rule mining method. In section V, we provide experiment and in section Ⅵ a conclusion.

2 The process of large-scale network behavior analysis and feature extraction

2.1 Overview

Basically, in this paper, the process of large-scale communication network behavior analysis and feature extraction consists of the following five steps:

Step1 Compute entropy of several flow level traffic features collected over each time bin.

Step2 Apply Principal Component Analysis and subspace method to entropy time series.

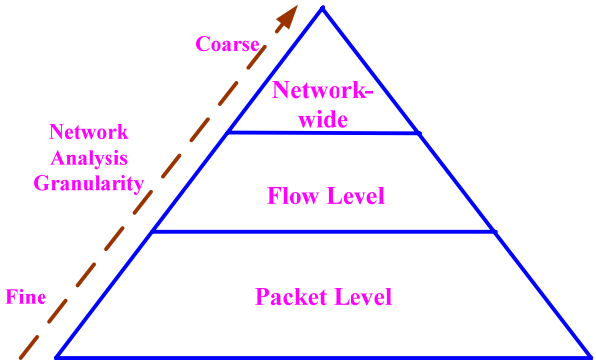


Figure 1 Large-scale communication network behavior analysis granularity

TABLE 1. DESCRIPTION OF SIX NETWORK SECURITY METRICS

Network security metrics	Description
H(srcPort)	Entropy of source port distribution
H(dstPort)	Entropy of destination port distribution
H(srcIP)	Entropy of source IP address distribution
H(dstIP)	Entropy of destination IP address distribution
H(octets)	Entropy of octets distribution
H(prot)	Entropy of protocol type distribution.

Step3 Apply time frequency analysis method and Piecewise Aggregate Approximation and Symbolic Aggregate approximation to anomaly time series.

Step4 Apply multiple motif pattern association rule mining to symbolic sequence.

Step5 Real-time monitor with valid motif association pattern.

The first 1,2,3 step is data preprocessing stage, step 4 is data mining stage, step 5 is the outcome of the mining and network traffic monitoring with the valid association rules

2.2 Large-scale Communication Network Behavior Analysis Granularity

The large-scale network behavior analysis consists of three levels, from bottom to top are packet level, flow level, network-wide level.

2.2.1 Packet Level Behavior Analysis

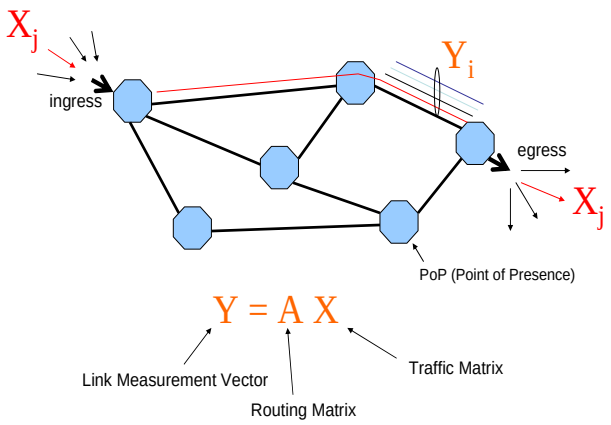


Figure 2 The demo of Network-wide Level Behavior Analysis

Each packet including time stamp, IP address or prefix, port number, protocol type, bytes and content. IP header information includes traffic volume by IP addresses or protocol, burst of the stream of packets, packet properties (e.g., sizes, out-of-order). TCP header information includes traffic breakdown by application (e.g., Web), TCP congestion and flow control, number of bytes and packets per session. Application header information includes URLs, HTTP headers (e.g., cacheable response?), DNS queries and responses, user key strokes and so on.

2.2.2 Flow Level Behavior Analysis

Basic information about the flow include source and destination IP address, port number, packet and byte counts, start and end times, ToS, TCP flags. Information related to routing includes next-hop IP address, source and destination AS.

2.2.3 Network-wide Level Behavior Analysis

Network-wide level traffic analysis combines traffic, topology, and state information. Network-wide level behavior analysis is mainly referred to traffic matrix analysis in this paper. Traffic Matrices (TM) reflect the traffic volume of OD (origin-destination) flow in a large-scale communication network. The demo of network-wide level behavior analysis is shown in Figure 2.

3 Data Preprocessing

3.1 Shannon Entropy and Flow Level Network Security Metrics

3.1.1 Shannon Entropy

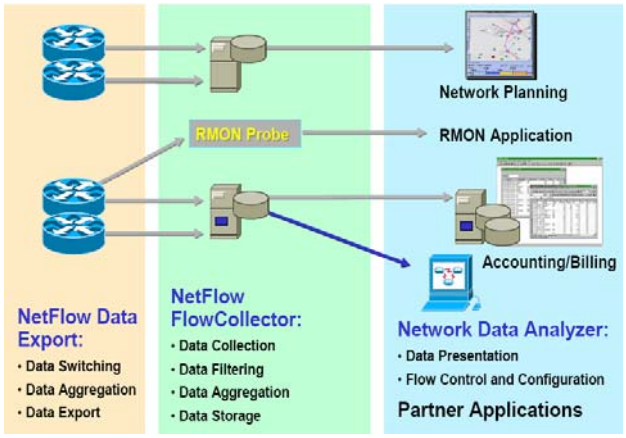


Figure 3 Cisco IOS NetFlow Infrastructure(Image From NetFlow PPT by Michael Lin, Cisco Systems)

Entropy is a metric that captures the degree of dispersal or concentration of a distribution, which is a measure of the uncertainty of a random variable. A wide variety of anomalies will impact the distribution of one of the discussed IP features. Let X be a discrete random variable with alphabet χ and probability mass function $p(x) = \Pr\{X = x\}, x \in \chi$, the *entropy* $H(X)$ of a discrete random variable X is defined by

$$H(X) = -\sum_{x \in \chi} p(x) \log p(x) \qquad (1)$$

$\Pr\{X = x\}$ is the probability of event occurring. For example, the probability of seeing IP 129.173.192.0 is defined to be number of packets using IP 129.173.192.0 divided by the total number of packets in the given time interval.

3.1.2 Flow Level Network Security Metrics

We can monitor applications and identify malicious traffic with *Netflow* information . The Cisco IOS NetFlow Infrastructure is shown in Figure 3. *Netflow* is passive monitoring tool include: Statistics about groups of related packets (e.g., same TCP/IP headers and close in time); Records header information, counts, and time. The flow record contents can be divided into two parts: One is the basic information about the flow; the other one is information related to routing. The basic information about the flow are as follows: the first information is source and destination, IP address and port; the second one is packet and byte counts; the third one is start and end times; the fourth one is ToS and TCP flags. The information related to routing are listed as follows: the first Next-hop IP address; Source and destination AS; Input and output.

Network security metrics refer to metrics are used for network security. In general, most of the anomalies affect the distribution of the six flow level network security metrics as follows: source address (sometimes called source IP and denoted srcIP), destination address (or destination IP, denoted dstIP), source port (srcPort), destination port (dstPort), octets and type of protocol, which are shown in Table 1.

3.2 Principal Components Analysis and Subspace Method

3.2.1 Principal Components Analysis

A fundamental unsupervised dimensionality reduction method is Principal Components Analysis (PCA). The basic idea of PCA can be illustrated as follows. On the one hand, PCA find the embedding subspace that gives the best approximation to the original samples; On the other hand, PCA is equivalent to finding the embedding subspace with the largest variance.

Let $S^{(t)}$ be the *total scatter matrix* :

$$S^{(t)} = \sum_{i=1}^{n}(x_i - \mu)(x_i - \mu)^T \qquad (2)$$

where $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$. The PCA transformation matrix T_{PCA} is defined as

$$T_{PCA} = \arg\max_{T \in R^{d\times r}}[tr(T^T S^{(t)} T(T^T T)^{-1})] \qquad (3)$$

3.2.2 Subspace Method

The main idea of subspace method in network behavior analysis is divide traffic space into normal subspace and abnormal space using PCA mentioned in section 3.2.1. For some m, the normal subspace S is the space spanned by PC_1 through PC_m , and the abnormal subspace $\tilde{S}$ is similarly the space spanned by PC_{m+1} through $PC_n (m \le n)$. All flow traffic at one time can be represented as in Figure 3

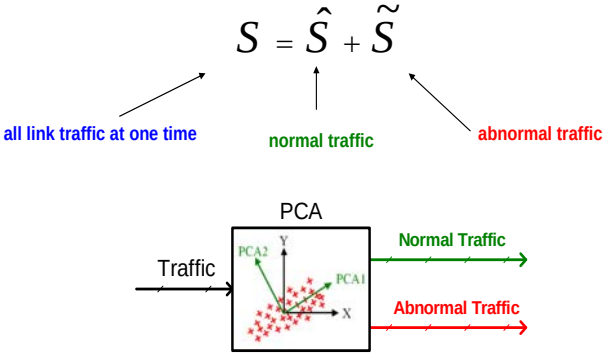


Figure 4 Subspace method

3.3 Time Series Representation and Time Frequency Analysis

3.3.1 Piecewise Aggregate Approximation

The basic idea of Piecewise Aggregate Approximation (PAA) [7][8][9] is that it represents the time series as a sequence of rectangle basis functions. Let n be the length of sequence, N be the number of PAA segments. $\overline{p_i}$ is the average value of the ith segment which can be computed by

$$\overline{p_i} = \frac{N}{n}\sum_{j=\frac{n}{N}(i-1)+1}^{\frac{n}{N}i} p_j \qquad (4)$$

The properties of PAA are that PAA approximates the original signal by a linear combination of N PAA coefficients for each frame and the PAA resulting transform is similar to Haar wavelet transform. Dimensionality reduction using PAA can reduce the signal from n to N dimensions and can guarantee no false dismissals with a special distance measure. The Advantages of PAA is: PAA supports flexible query lengths and weighted Euclidean distance; PAA is easy to implement and faster than DFT (the time complexity of PAA is $O(nm)$ where m is the number of frames).

3.3.2 Symbolic Aggregate approximation

The basic idea of Symbolic Aggregate approximation (SAX)[7][8][9]is that it converts the time series into an discrete symbolic sequence. Having transformed a time series data into the PAA,

TABLE 2. A LOOKUP TABLE THAT CONTAINS THE BREAKPOINTS

β_i \ c	3	4	5	6	7	8
β_1	-0.43	-0.67	-0.84	-0.97	-1.07	-1.15
β_2	0.43	0	-0.25	-0.43	-0.57	-0.67
β_3		0.67	0.25	0	-0.18	-0.32
β_4			0.84	0.43	0.18	0
β_5				0.97	0.57	0.32
β_6					1.07	0.67
β_7						1.15

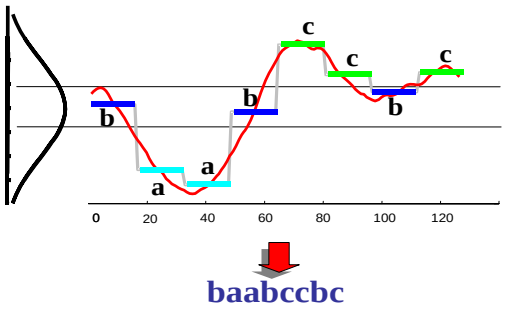


Figure 5 The demo of SAX method

we can apply SAX to obtain a discrete symbolic representation. Since normalized time series have a Gaussian distribution, we can determine the "breakpoints" that will produce c equal-sized areas under Gaussian distribution curve [10].

The breakpoints will be found out by looking them up in Table 2. Once the breakpoints have been obtained we can disperse time series into discrete symbolic series. We firstly obtain a PAA manner of the origin time series. All PAA coefficients that are below the smallest breakpoint are transformed to the symbol "a", and all coefficients greater than or equal to the smallest breakpoint but less than the second smallest breakpoint are transformed to the symbol "b", etc. Figure 5 illustrates the idea.

3.3.3 Instantaneous Frequency

Because single component signal has only one frequency at any time, Instantaneous Frequency(IF) can be used to describe local frequency behavior. For any real value signal $x(t)$, we define analytic signal $x_a(t)$ associated to $x(t)$ as in (5):

$$x_a(t) = x(t) + jHT(x(t)) \tag{5}$$

$HT(x)$ is the Hilbert Transform of x .From this analytic signal, we can define the concepts of instantaneous frequency as in (6):

$$f(t) = \frac{1}{2\pi}\frac{d\arg x_a(t)}{dt} \tag{6}$$

Based on the description of instantaneous frequency, we can capture the local behavior of the signal.

3.3.4 Wavelets Transform and Wavelet Packet Transform

Suppose $x(t) \in L^2(R), \psi(t)$ is the basic wavelet function, and $\psi_{a\tau}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-\tau}{a})$ is the shift and scale extension of the basic wavelet function as in

$$WT_x(a,\tau) = \frac{1}{\sqrt{a}}\int x(t)\psi^*(\frac{t-\tau}{a})dt = <x(t),\psi_{a\tau}(t)> \tag{7}$$

is called as the wavelet transform of $x(t)$. The basic idea of discrete wavelet transform (DWT) is to transform a discrete time signal into a discrete wavelet representation. Discrete wavelet transform converts an input series $x_0, x_1, ..., x_k$ into one high-pass wavelet coefficient series and one low-pass wavelet coefficient series (of length $n/2$ each) given by as in (8) and (9):

$$H_i = \sum_{k=0}^{m-1} x_{2i-k} \bullet s_k(z) \tag{8}$$

$$L_i = \sum_{k=0}^{m-1} x_{2i-k} \bullet t_k(z) \tag{9}$$

Where $s_k(z)$ and $t_k(z)$ are wavelet filters, m is the length of the filter, and $i = 0,1,...[n/2]-1$. In practice, such transformation will be used recursively on the low-pass series until the desired number of iterations is reached.

Wavelet Packet Transform (WPT) is a method which makes time frequency decomposition of signal. WPT is of self-adaptive of signal, which can effectively display the time frequency property of signal. Just by orthogonal mirror filter we can obtain WPT decomposition. Assume the signal $y(t)$, we can obtain

$$\begin{cases} y_{2n}(t) = \sqrt{2}\sum_k h(k) y_n(2t-k) \\ y_{2n+1}(t) = \sqrt{2}\sum_k g(k) y_n(2t-k) \end{cases} \quad (10)$$

Function set $\{y_n(t)\}$ is called as wavelet packet which is the result of whole decomposition of all bands on various scale of origin signal. Let $k = n - 2^j$, then $y_n(t) = y_{2^j+k}(t)$ is the result of decomposition of k -band on scale j. WPT may consist of various orthogonal basis, wavelet basis is the typical case. Among all combination, the least entropy is the good basis. The decomposition of good basis can represent the time-frequency of signal which imply that the method is adaptive for signal.

3.3.5 Choi-Williams Distribution

In order to reduce the disturbed components of the Chio-William distribution [11], we should have a research on the factors which make the disturbed value minimum. The Choi-Williams is proposed to solve this problem as in

$$C(t,f) = \int_{-\infty}^{+\infty} e^{j2\pi ft} \int_{-\infty}^{+\infty} \sqrt{\sigma/4\pi\tau^2}\, e^{\frac{\sigma(\mu-t)^2}{4\tau^2}} x(\mu+\frac{\tau}{2}) x^*(\mu-\frac{\tau}{2}) d\mu d\tau \quad (11)$$

Wigner-Ville distribution $C(t, f)$ satisfies the edge conditions and shift characteristics but does not satisfy weak and limited support characteristics. However,when $\sigma \to \infty$ it would satisfy weak and limited support characteristics.

3.3.6 Pseudo Wigner-Ville Distribution

For a given time, Wigner-Ville Distribution can describe the global distribution of a signal. In addition, for a given frequency, it can also equally measure all the frequencies either higher or lower than the given frequency. In fact, we are not able to study all the integrals between $-\infty$ and $+\infty$ but the ones in a limited range. When studying the distribution shape of a certain time (t) we should study the features of signals near the given time. In a sense, it means to condense the cross-item of multivariable signal by adding some windows and deleting the non local components. Finally we change Wigner-Ville distribution into local distribution. Pseudo Wigner-Ville distribution characterizes [11]a local behavior of a signal as the following formula, so it is convenient for us to mine the local features of the fault signals as in

$$PW_x(t, f) = \int_{-\infty}^{+\infty} h(\tau) x(t+\frac{\tau}{2}) x^*(t-\frac{\tau}{2}) e^{-j2\pi f\tau} d\tau \quad (12)$$

$h(t)$ is the window function.

4 Multiple Motif Pattern Association Rule Mining

4.1 Apriori

Apriori is a far-reaching algorithm proposed by R.Agrawal and R.Srikant [13][14][15] in 1994 for mining frequent itemsets for Boolean association rules. The name of the algorithm is based on the fact that the algorithm uses priorknowledge of frequent itemsets properties. The Apriori property is that all nonempty subsets of a frequent itemset must also be frequent. Algorithm 1 shows the pseudo-code for Apriori algorithm. Step 1 of Apriori finds the frequent 1-itemsets L_1 . In step 2 to step 11, L_{k-1} is used to generate candidate C_k in order to find L_k for $k \geq 2$ .The apriori_gen procedure generates the candidates and then uses the Apriori property to eliminate those having a subset that is not frequent (step 3).This procedure is described below. Once all of the candidates have been generated,the database is scanned (step 4). For each transaction, a subset function is used to find all subsets of the transaction that are candidates (step 5), and the count for each these candidates is accumulated (steps 6 and 7). Finally, all of those candidates satisfying minimum support (step 10) form the set of frequent itemsets, L (step 12).

Input: D, a database of transactions;
min_sup, the minimum support count threshold.
Output: L, frequent itemsets in D.
Method:

1 L_1=find_frequent_1-itemsets(D);
2 **for** k=2; $L_{k-1} \neq \phi; k++$ **do**
3 C_k=**apriori_gen**(L_{k-1});
4 **foreach** *transaction* $t \in D$ **do**

```
5      C_t =subset(C_k , t);
6      foreach candidate c ∈ C_t do
7        c.count++;
8      end
9    end
10   L_k = {c ∈ C_k | c.count ≥ min_sup}
11 end
12 return L = ⋃_k L_k ;
```

Algorithm 1: The Apriori Algorithm

The apriori_gen procedure performs two kinds of actions, namely, **join** and **prune**, as described above. In the join component, L_{k-1} is joined with L_{k-1} to generate potential candidates(steps 2 to 5).The prune component (steps 6 to 7) employs the Apriori property to remove candidates that have a subset that is not frequent. The test for infrequent subsets is shown in procedure has_infequent_subset.

```
Procedure apriori_gen( L_{k-1} :frequent(k-1)-itemsets)
1  foreach itemset l_1 ∈ L_{k-1} do
2    foreach itemset l_2 ∈ L_{k-1} do
3      if
         (l_1[1] = l_2[1]) ∧ (l_1[2] = l_2[2]) ∧ ...
         (l_1[k-2] = l_2[k-2]) ∧ (l_1[k-1] < l_2[k-1])
       then
4        c = l_1 ⋈ l_2 ;
5        if has_infrequent_subset (c, L_{k-1} ) then
6          delete c ;
7        else add c to C_k ;
8      end
9    end
10   return C_k ;
11 end
```

```
1 Procedure has_infrequent_subset( c :candidate
    k-itemset; L_{k-1} :frequent(k-1)-itemsets)}
2 foreach (k-1)-subset s of c
3   if s ∉ L_{k-1} then
4     return TRUE;
5   end
6   return FALSE;
7 end
```

Let $I = \{I_1, I_2 ..., I_m\}$ is a set of items. X is an itemset if it is a subset of I. $D = \{t_i, t_{i+1}, ..., t_n\}$ is a set of transaction. A transaction t contains itemset X if and only if, for all items, where $i \in X, i$ is a

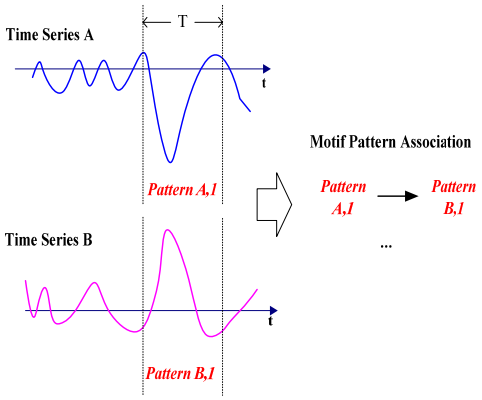


Figure 6 Multiple motifs association rules data mining at the same time interval T

$t-$itemset. An itemset X in a transaction database D has a support, denoted as $Supp(X)$,see as in (13).

$$Supp(X) = \frac{|X(t)|}{|D|} \tag{13}$$

where $X(t) = \{t \text{ in } D \mid t \text{ contains } X\}$.

The support of a rule $X \to Y$ is $Supp(X \to Y)$ as the support of $X \bigcup Y$, the confidence of a rule $X \to Y$ is $Conf(X \to Y)$ as the ratio $Supp(X \bigcup Y) / Supp(X)$.

4.2 Multiple Motif Pattern Association Rule Mining

These six network security metrics mentioned above, collected over each time bin, make up a six-dimensional time series. This time series is put into the data matrix X, where features vary across columns of X and time varies across rows. For a bin size of 5 minutes and 8 day-long trace, X is thus 2304×6.

Motif discovery in multiple time series is an important problem with great significance. A dependency is an unexpectedly frequent or infrequent co-occurrence of anomalies over time. This indicates that an anomaly motif on one time series is related to other anomalies motif on other time series, which seems to be independent from the former abnormal motifs patterns. For example, rise and fall of entropy value on some anomaly detection metrics obviously cause entropy value of one network security metrics to rise and fall. If we analyze the multiple time series for some network security metrics and we can discover dependencies

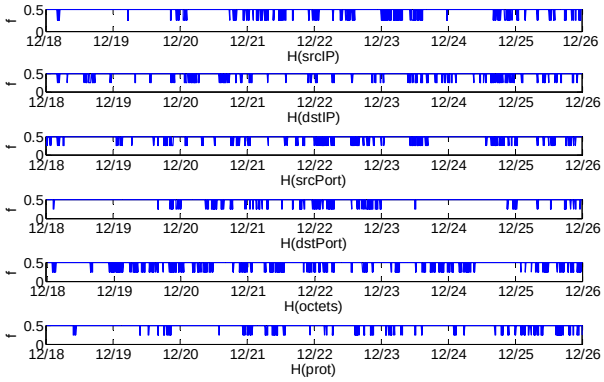
between all network security metrics, and the dependencies can help us to decide better time to monitor network security situation. So, the dependencies can be expressed as rules. In our case,

Figure 7 Abilene Network

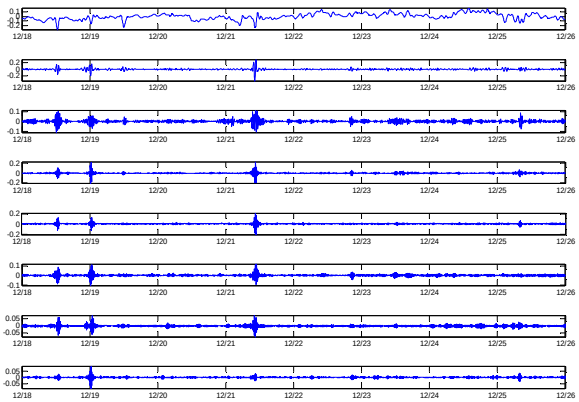
these dependencies are called motif association rules. Strong dependencies capture structure in the network traffic flows because it indicates that there is relationship between their constituent patterns that is occurrences of those patterns are not independent. The association rule discovery problem usually translates into finding all sets of patterns of dependencies that satisfy a pre-specified minimum threshold on support, and then post-processing them to find the interesting rules. Such dependencies are called frequent. The association rules usually predict the occurrence of some other set of dependencies with certain degree of confidence. The motif association rules, which are the results of multiple time series association rules mining, lay the foundation for us to analyze anomalous behavior of large-scale communication network which is shown in Figure 6. Algorithm 2 shows the pseudo-code for Multiple Motif Pattern Association Rule Mining algorithm.

Input: *D*, a database of Cisco Netflow data;
min_sup, the minimum support count threshold.
Output: *R*, anomaly motifs rules;
T, the time distribution of the anomaly motifs rules.
Method:
EntropyTS ← **ComputeEntropy**(*D*);
AnomalyTS ← **PCA**(EntropyTS);
SymbolicTS ← **SymbolicRepresent**(*AnomalyTS*);
AnomalyMotifs ← **FindAnomaly**(*AnomalyTS*);
SymbolicAnomalyMotif ← **SymbolicRepresent**(*AnomalyMotifs*);
Rules ← **Apriori**(*SymbolicTS*);
R ← **Match**(*SymbolicAnomalyMotif* , *Rules*);
T ← **TimeDistribution**(R);

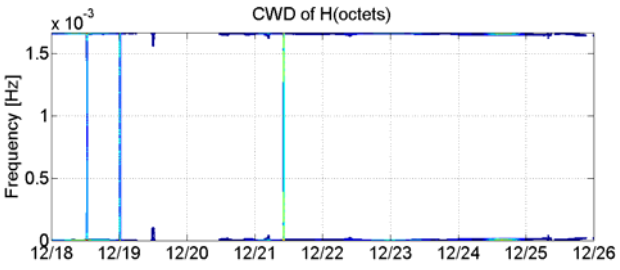
Algorithm 2: Multiple Motif Pattern Association Rule Mining



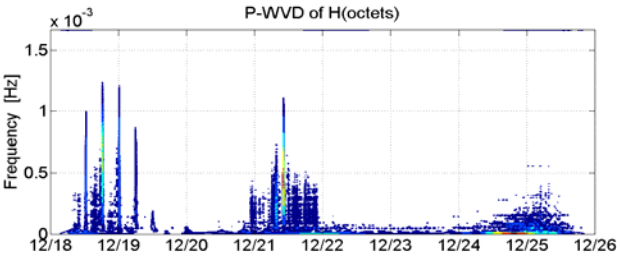
(a) Apply IF to six network security metrics



(b) Apply WPT to H(octets)



(c) Apply CWD to H(octets)



(d) Apply PWVD to H(octets)

Figure 9 Apply IF,WT,CWD and PWVD methods to anomaly entropy time series

5 Experimental Results

5.1 Data Sets

The sampled flow data collected from the backbone networks: Abilene [1]. Abilene is the Internet2 backbone network, connecting over 200 US universities and peering with research networks in Europe and Asia. It consists of 11 Points of Presence (PoPs), spanning the continental US. Sampling is periodic, at a rate of 1 out of 100 packets, which is shown in Figure 7.

5.2 Simulation Test

We collect data from 08:05 a.m. on December 18th, 2006 to 08:00 a.m. on December 26th, 2006.We firstly compute entropy value of each network security metrics within each time bin. We analyze six time series of entropy value by PCA method, and obtain the time series of the anomalies' entropy value.

In order to locate the anomalies, we apply the Instantaneous Frequency, Wavelet Packet Transformation, Choi-Williams distribution and Pseudo Wigner-Ville distribution methods to anomaly entropy time series. The results are shown in Figure 8. From the Figure 8, we can make sure some anomalies （worms）certainly exist in 20:40 Dec.18.

In order to discover the patterns of anomalous motifs, (1)we apply PAA and SAX to six network security metrics data, (2) according to the anomaly time points obtained from anomaly detection phrase, we extract the anomaly motifs (i.e. anomaly motif between 146 point and 155 point), the outcome is shown by Figure 9, (3) we should analyze the rule of all six network security metrics data which satisfy the known anomaly motif pattern resulted from (2) and how often this rule happen with association rules mining. The step (3) can be described as follows:

Let $H(srcIP)=1, H(srcPort)=2, H(dstPort)=3, H(Octets)=4, H(prot)=5, H(dstIP)=6$, 'A' denotes the lowest entropy value, 'B' denotes the lower entropy value, 'C' denotes the medium entropy value, 'D' denotes larger entropy value, 'E' denotes the largest entropy value. Then we apply association rules mining to

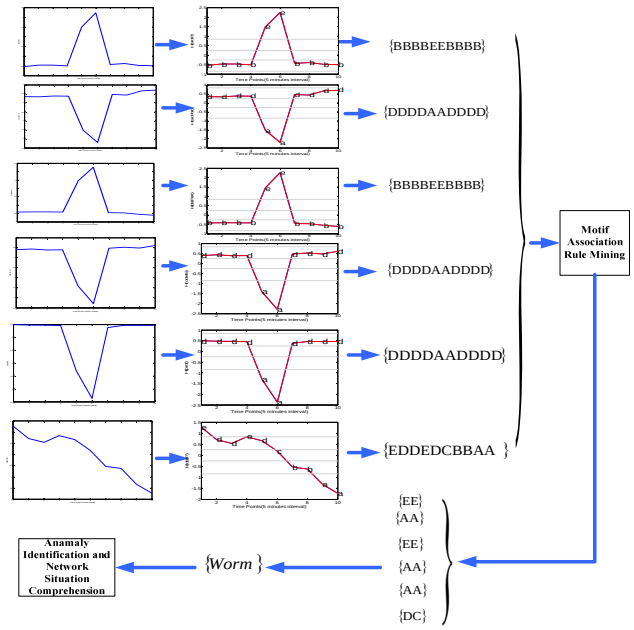


Figure 9 Apply SAX on 6 time series

TABLE 3. THE DISTRIBUTION OF "$1EE,2AA,4AA,5AA \Rightarrow 3EE$"

Day	18	19	19	19	19	19	19	19	19	21	25
Hour	20	05	05	06	08	20	20	20	20	18	16
Minutes	40	30	40	40	40	00	10	20	30	20	00

alphabet sequence of eight days (from Dec.18 to Dec.25) to get the association rules of the anomaly pattern in the backbone IPLSng router:

$1EE, 2AA, 4AA, 5AA \Rightarrow 3EE$

$(\sup = 11, conf = 100)$

The rule $1EE, 2AA, 4AA, 5AA \Rightarrow 3EE$ means that the distribution of source address is dispersive ($E \to E$, the curve doesn't increase and doesn't decrease), the distribution of source port is concentrated, the distribution of octets is concentrated, and the distribution of protocol is dispersive is also concentrated, from which we can infer that the distribution of destination port is dispersive. Hence this rule can be used to identify whether the anomaly is worm or not. The experimental result is shown in Table 3.

6 Conclusion

In this paper, firstly, we compress Netflow data using Shannon entropy; secondly, we use PCA for dimensional reduction; thirdly, we apply the wavelet packet transformation, Choi-Williams distribution and Pseudo Wigner-Ville distribution method to locate anomaly motifs; fourthly, we apply PAA and SAX to six network security metrics data to obtain six discrete symbolic sequence and extract the anomaly motifs ; finally, we analyze the rule of all six network security metrics data which satisfy the known anomaly motif pattern resulted from previous step and how often this rule happen with association rules mining method. From experimental results, we can arrive at the conclusion that our method can be used to analyze large-scale communication network behavior.

We evaluate the method on six network security metrics anomalies, which are specific instance of flow level behavior outlier resulting from unusual changes in the flow traffic. We showed how to use our method to locate outliers from simple and readily available flow measurement. We quantified the efficacy of our method on *Netflow* data collected from backbone networks.

Our ongoing work is centered on the improvement of the analysis precision of network behavior anomalies and the problem of high dimensional data.

7 Acknowledgments

The authors would like to acknowledge the Netflow data of Abilene and thank reviewers for their helpful comments. This research is supported in part by the Natural Science Foundation of China under grant 60872033, Program for New Century Excellent Talents in University under grant NCET-07-0148 and Major National Basic Research Development Program of China (“973” Program, No. 2007CB307100)

References:

[1] http://abilene.internet2.edu.

[2] http://www.snort.org

[3] Ke Wang, Salvatore J.Stolfo, “Anomalous Payload-based Network Intrusion Detection”, *the 7th International Symposium on Recent Advances in Intrusion Detection*,2004

[4] Masaki Ishiguro, Hironobu Suzuki, Ichiro Murase, Hiroyuki, “Internet Threat Detection System Using Bayesian Estimation,” *the 16th Annul FIRST Conference on Computer Security Incident Handling*,2004

[5] Lakhina,A.,Crovella,M.,and Diot,C,“Mining anomalies using traffic feature distributions”. In *ACM SIGCOMM*, Philadelphia, Pennsylvania, USA, 2005),pp.217–228.

[6] Hotelling,H,“Analysis of a complex of statistical variables into principal components,”*J. Educ. Psy*.(1933), 417–441.

[7] Lin,J.,Keogh,E., Lonardi, S. & Chiu, B, “A Symbolic Representation of Time Series, with Implications for Streaming Algorithms,” In *proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, San Diego, CA. June 13, 2003

[8] Lin,J.,Keogh,E., Patel, P. & Lonardi, S, “Finding Motifs in Time Series". In *proceedings of the 2nd Workshop on Temporal Data Mining, at the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Edmonton, Alberta, Canada. July 23-26, 2002

[9] Eamonn J. Keogh ,Michael J. Pazzani, “An Enhanced Representation of Time Series Which Allows Fast and Accurate Classification,Clustering and Relevance Feedback. In *International Conference on Knowledge Discovery and Data Mining*, pages 239–243, New York,NY, USA, August 1998.

[10] R.J.Larsen and M.L.Marx. *An Introduction to Mathematical Statistics and Its Applications*.2nd ed. Englewood,Cliffs,NJ:Prentice Hall.1986

[11] L.Cohen,*Time-Frequency Analysis:Theory and Applications*. Prentice Hall.1998

[12] R.Agrawal, T. Imielinski, A. Swami, “Mining association rules between sets of items in large databases,” in *Proc. ACM SIGMOD Int. Conf. Management Data*, 1993, pp. 207–216.

[13] R.Agrawal;T.Imielinski;A.Swami, “Mining association rules between sets of items in large databases,” in *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*, Jun. 1993.

[14] Chengqi Zhang, Shichao Zhang. *Association Rule Mining:models and algorithms*.2002

[15] J. Han; M. Kamber, *Data Mining:Concepts and Techniques*, San Francisco: Morgan Kaufmann, 2006.