Anomaly Detection of Network Traffic Based on Wavelet Packet

Jun Gao Guangmin Hu Xingmiao Yao
Key Lab of Broadband Optical Fiber Transmission and Communication Networks,
UESTC, Chengdu 610054, China
{jgao,hgm,yxm}@uestc.edu.cn

Rocky K. C. Chang
Department of Computing
The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong
{csrchang}@comp.polyu.edu.hk

***Abstract*— The rapid and accurate detection of network traffic anomaly is one of the preconditions to guarantee the effective work of the network.Aiming at the deficiency of present methods of network traffic anomaly detection, we propose a scale-adaptive method based on wavelet packet. By means of wavelet packet decomposition, our method can adjust the decomposition process adaptively, has the same detective ability to the anomaly of various frequency, especially the middle and high frequency ones which can not be checked out by the multi-resolution analysis. By means of adaptive reconstruction of the wavelet packet coefficient of different wavelet domains which anomaly, our method is able to confirm the characteristics of anomaly and enhance the reliability of detection. By means of a fast wavelet packet algorithm based on sliding window, our method satisfies can decrease the computation complexity of wavelet packet transform. By means of scale-adaptive detection window choice method based on wavelet center frequency, we can choose different detection time-windows to anomalous traffic which has difference scale. The simulation results prove that the method can detect the network traffic anomaly efficiently and rapidly.**

I. Introduction

The network traffic anomaly refers to the status that traffic behaviors depart from the normal behaviors. Many reasons, such as the misuse of network equipments, network operations anomaly, flash crowd, network intrusion and so on will cause network traffic anomaly. The characteristic of anomaly traffic is that it breaks out without any omen and can destroy networks and computers in a short time (For instance, the outburst of traffic behavior caused by specific attack programs or worm burst). Therefore, detecting traffic anomaly rapidly and accurately is one of the preconditions of ensuring the efficient network operation. The traffic in routers especially the routers in trunk network is very large and change continually, while the anomalous traffic is small, compared with the normal traffic and changes of the normal traffic. The ultimate aim of the anomaly detection algorithm is to detect the relatively small anomaly traffic from the relatively large background traffic. It is very difficult to implement this aims, so detection of anomalous traffic has become the attractive and valuable subject in the present academic and industrial circles.

Various schemes are proposed for the network traffic anomaly detection, such as the case based reasoning approach[1][2], the limit state machine approach[3][4], mode matching approach[5][6], statistical analysis approach[7][8], Hurst parameter analysis approach[9] and subspace analysis approach[10][11][12], etc. The achievements of former researches have greatly promoted the development of the anomaly detection and improved the detection results constantly. However, due to the complexity of network traffic anomaly detection, there were problems in the real-time performance and the accuracy of detection still. Researchers have found that almost almost all the traffic time-varying signal were multi-scales [13], and the time-varying signals of the normal network traffics and that of the abnormal network traffics were different in frequency band range. That is to say, the difference between anomalous traffics and background traffics is various in different frequency bands. In certain frequency band, the energy of anomalous traffics is rather high in proportion to the total energy, so anomaly detection could be done easily. The wavelet transform can get arbitrary signal characteristic of time-frequency domain, which can help to explore the transient abnormal phenomenon from normal signals and demonstrate its components. Therefore, researchers put forward the wavelet analysis approach[14][15][16][17][18].

As a new technology, anomaly detection of network traffic based on wavelet analysis is taken more seriously recently years. V.Alarcon-Aquino presented an algorithm based on undecimated discrete wavelet transform and bayesian analysis[14]. This algorithm is able to detect and locate subtle changes in variance and frequency in the given time series, but its decomposition scale is limited and the algorithm is complicated. Anu Ramanathan presented a WADeS (Wavelet based Attack Detection Signatures) mechanism based on wavelet analysis to detect the DDoS attack[15], which makes wavelet transform for the traffic signals, then computes the variance of the wavelet coefficients to estimate the attack points. However, this method has very high computation complexity. Barford presented a method[16] presented a method which decomposes network traffic with decimated discrete wavelet transform , then synthesizes to Low, Mid, High frequency-parts, and finally detect anomaly with Deviation Scoring respectively. This algorithm is able to detect the flash crowds and short-term anomalies in postmortem. But it doesn't solve the problems of adaptive choice of detective scale and detection time-windows,has high computation complexity. Seong Soo Kim proposed a technique for traffic anomaly detection based on

analyzing correlation of destination IP addresses in outgoing traffic at an egress router[17]. This technique can be employed for postmortem and real-time analysis of outgoing network traffic, but it has not the same detective ability towards various frequency anomaly since it is based on multi-resolution analysis. Lan Li proposes an energy distribution based on wavelet analysis to detect the DDoS attack[18]. Research finds the energy distribution variance changes markedly causing a "spike" when traffic behaviors affected by DDoS Attack.

Generally speaking, four problems exist in the present anomaly detection methods based on wavelet transform: (1)Almost all the algorithms use multi-resolution analysis, so the poor resolution in middle-high frequency leads to the insufficiency of comprehensiveness in anomaly detection. That is to say, the detection method only works well on low frequency anomaly. (2)An anomaly probably distributes in many discontiguous frequency bands, so the detection is unauthentic if only in one scale. (3)Wavelet transform has very high computation complexity, and the real-time performance of anomaly detection can't be guaranteed because of lack of fast algorithm. (4)It is difficult to decide the time-windows of anomaly detection; usually the time-windows are the same in each scale, and there are no corresponding windows selected according to anomalous signals themselves. To solve the problems described above, this paper proposes a new network anomaly detection method based on wavelet packet transform. our method can adjust the decomposition process adaptively, and has the same detective ability to middle and high frequency as well as low frequency anomaly. By means of adaptive reconstruction of the wavelet packet coefficient of different wavelet domains which anomaly, our method is able to confirm the characteristics of anomaly and enhance the reliability of detection. By means of a fast wavelet packet algorithm based on sliding window, our method reduce computation complexity of wavelet packet transform remarkably. By means of scale-adaptive detection window choice method based on wavelet center frequency, our method can choose different detection time-windows to difference scale's anomalous traffic.

The rest of paper is organized as follows: In Section II we describe the wavelet packet transform and its fast algorithm. In Section III we describe our method for anomaly detection. In Section IV we illustrate the simulation results of our detection method, and Section V concludes the paper.

II. Wavelet Packet Transform And Its Fast Algorithm

Generally speaking, the energy of power spectral density (PSD) of the normal traffic in each frequency band is relatively well-proportioned, but that of PSD of the anomalous traffic is concentrated in certain frequency bands. Researchers used wavelet analysis to detect anomaly just based on the differences between the normal and anomalous traffic signals in the frequency domain. Since the scale of the present detection algorithms based on multi-resolution analysis is binary; the frequency resolution is relatively poor in high frequency. However, various causes of traffic anomalies lead that anomalies may occur in low frequency or in high frequency. Therefore, their approaches cannot detect all of frequency anomaly traffic effectively.

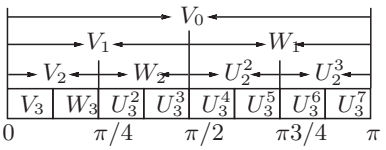


Fig. 1. Frequency domain division of WPD at level 3

Wavelet packet analysis which can provide a more accurate method for signal, divides the frequency bands into various levels, further decomposes high-frequency section which cannot be decomposed by multi-resolution analysis. And according to the characteristics of analyzed signal, it can self-adapt to choose corresponding bands, to match the frequency spectrum of signal, so to improves time-frequency resolution.

A. Wavelet Packet Analysis

Wavelet transform representing signals with wavelet coefficients and wavelet series is composed by translation and dilation of mother wavelets. Dyadic wavelet transform, because of its dyadic scale variety, has not so fine frequency resolution as wavelet packet decomposition which divides not only the low frequency part of signal but also the high frequency part which wavelet transform does not. The frequency domain of the signals is divided finer and finer, as shown in Figure 1 where $V_i(i = 1, 2, 3), W_1, W_2$ and $U_i^j(i = 2, 3; j = 2, 3, \ldots, 7)$ are frequency domains at different levels; and V_0 is the entire domain.

The algorithms of wavelet packet decomposition (WPD) are

$$x_{2m}^j(n) = \sum_k h(k - 2n)x_m^{j-1}(k), \tag{1}$$

$$x_{2m+1}^j(n) = \sum_k g(k - 2n)x_m^{j-1}(k). \tag{2}$$

The reconstruction algorithm of WPD is

$$x_m^{j-1}(n) = \sum_k \bar{h}(n - 2k)x_{2m}^j(k) + \sum_k \bar{g}(n - 2k)x_{2m+1}^j(k) \tag{3}$$

where h and g are the filter coefficients; k is equal to 1,2,3...; $x_m^j(n)$ is the m-th sequence through j-th level WPD of $f(n)$; $\bar{h}$ and $\bar{g}$ are respectively the mate operators of h and g.

If decomposition level is N and the highest frequency of original signals is f_h, then the width of each frequency band corresponding to wavelet packet tree node is $f_h/2^N$. The frequency division procedure of WPD can be represented as

$$V_0 = \bigoplus_{j \in Z} W_j \tag{4}$$

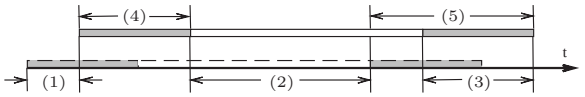


Fig. 2. A real-time algorithm of wavelet packets transform.(1) The moving points ΔN,(2) the length of the same data ragion,(3) The length of the data beyond the boundary

B. A Fast Wavelet Packet Algorithm Based On Sliding Window

Network traffic anomaly detection requires on-line work, wavelet transform requires a large amount of computation, thus it is hard to meet the need of real-time detection. In our researches, a sliding window needs to be built during the anomaly detection of network traffic, and before and after the sliding there are a lot of redundant data. Wavelet transform compute the redundant data twice, which take too much time to operate. By storing the part of wavelet packet coefficients in memory in advance, we avoid repeating calculation and gain a fast wavelet algorithm at the cost of increasing some storage space. The fast algorithm can sufficiently support our network traffic anomaly detection mechanism.

To make it easy, we choose rectangular window as the sliding window, N as the width of sliding window and the formula as follow:

$$w(n) = rect(n) = \begin{cases} 1, & -N/2 \le n \le N/2. \\ 0, & otherwise. \end{cases} \tag{5}$$

Its frequency spectrum is a $sinc$ function, as formula 6, where C is constant. So it has effect on the original signal.

$$w(j\omega) = Csinc(\omega/N) \tag{6}$$

Assume $x(n)$ as an infinitely long traffic signal serial and use window function $w(n)$ to intercept $x(n)$, we get:

$$s(n) = x(n)w(n) \qquad n = 0, 1, \ldots, N-1 \tag{7}$$

Assume ΔT as the moving points between two slide, and f_s as the sampling frequency, then the moving points between two slide is:

$$\Delta N = \Delta T \cdot f_s \tag{8}$$

Therefore, when $\Delta N < N$, there would be a lot of redundant data before and after the sliding. The smaller the ΔN , the more the redundant data. When the sliding window intercepts the signal $s(n)$, which is passing through the network traffic anomaly detector, the back-end wavelet transform would compute the redundant data twice, which takes too much time to operate and would affect a decline in real-time performance of the detector.

In the above deduction, we assume that the serial is infinite. But the serial is finite in practice. In order to meet the demand of the wavelet analysis, we have to use some appropriate data to extend the finite serial. We adopt the data periodic repetition method so as to reduce the edge effect. First of all, we will discuss the fast algorithm of wavelet based on $Mallat$ algorithm.

Set S_1 as the intercepted data before the window moves while S_2 as those after it moves; $f(n)$ as the filter function; $2L$ as the filter length. If j is the levels of wavelet packet decomposition, $N_j = N/2^j$ would be the wavelet packet coefficients length form the j-th level, and $\Delta N_j = \Delta N/2^j$ would be the moving points from the j-th level. The scaling coefficients of the j-th level are S_1^j and S_2^j. So the wavelet coefficients is:

$$DWT_1(k) = \sum_{i=-L}^{L} S_1^j(k-i)f(i) \qquad k = 0, 1, \ldots, N_j \tag{9}$$

Form formula 9 and Figure 2 we get:

1. When $k - i < 0$, we get $0 \le k < L$, the data in this section are out of range and demand process separately. This section is in the beginning of the signal.

2. When $k - i > N_j$, we get $N_j - L < k \le N_j$, the data in this section are out of range and demand process separately. This section is in the end of the signal.

3. When $L \le k \le N_j - L$, the data in this section do the normal convolution.

In the lapse of ΔT, the sliding window sample the traffic data to get the signal S_2. So we get:

$$S_2^j(k) = \begin{cases} S_1^j(k - \Delta N_j), 0 \le k \le (N_j - \Delta N_j). \\ newinputdata, (N_j - \Delta N_j) < k \le N_j. \end{cases} \tag{10}$$

The wavelet coefficients $DWT_1^j(k)$ of S_1^j in the section of

$$L + \Delta N_j \le k \le N_j - L \tag{11}$$

are the same as the wavelet coefficients $DWT_2^j(k)$ of S_2^j in the section of

$$L \le k \le N_j - L - \Delta N_j \tag{12}$$

Therefore, the same points of the $DWT_1^j(k)$ and $DWT_2^j(k)$ are:

$$M = N_j - 2L - \Delta N_j \tag{13}$$

The number of the different data points in the head is:

$$N_1 = L \tag{14}$$

The number of the different data points in the tail is:

$$N_2 = L + \Delta N_j \tag{15}$$

Therefore, as long as $N > L$, we only need to compute the wavelet coefficients of S_2 in the section of $k \in [0, L] \cup [N_j - (L + \Delta N_j), N_j]$. The wavelet packet fast calculation includes 4 steps:

1. Get the coefficients $DWT_1^j(k)$, after the wavelet transform on the signals S_1, and from formulas 11,12,14 store $k \in [L, N_j - L - \Delta N_j]$ section wavelet coefficients, which is represented in Figure 2(2).

2. According to the principle of data periodic repetition which deals with data beyond the boundary, and from formulas 13,15, we can pick up the $k \in [N_j - L - \Delta N_j, N_j] \cup [0, L + \Delta N_j]$ sections for signals S_2, and form signal series which convolute with filter to get the wavelet coefficients $DWT_2^j(k)$ of $k \in [0, L + \Delta N_j]$ section, as it is shown in Figure 2(4).

3. Make up signal series from the $[N_j - (L + \Delta N_j), N_j]$ and $[0, L + \Delta N_j]$ of signal S_2, and convolute with filter, we can get the wavelet coefficients $DWT_2^j(k)$ of $k \in [N_j - (L + \Delta N_j), N_j]$, as it is shown in Figure 2(5).

4. Synthesizing three steps mentioned above, we can get the wavelet packets transform coefficient of $DWT_2^j(k)$.

The fast algorithm of wavelet transform based on sliding window technology adds an array which is $M(M = \sum_j N_j - 2L - \Delta N_j)$ long, and used for recording the same wavelet coefficient of $DWT_1^j(k)$ and $DWT_2^j(k)$, by comparison with normal wavelet transform algorithm. Replace the $DWT_1^j(k)$ and $DWT_2^j(k)$ to wavelet packet coefficient $WPT_1^j(k)$ and $WPT_2^j(k)$, we can get the fast algorithm of wavelet packet transform.

III. Anomaly Detection of Network Traffic Method Based on Wavelet Packet

A. A Statistical Detection Algorithm

The statistical detection algorithm is improved by the deviation scoring algorithm[16] proposed by P. Barford. Two detection windows are applied to the deviation scoring algorithm, one is $HisWin$, the other $DetWin$, as it is shown in Figure 3. Both slide with time, and are for real-time update. At time t, we compute the variance V_1 in the detection window $(t - DetWin, t)$, and the variance V_2 in the historical window $(t - HisWin, t)$. We set $ratio = V_1/V_2$, then, to some extent, the parameter $ratio$ represents the deviation of the sample variance in the detection window compared to the historical normal data variance. If the traffic is anomalous in the detection window, there must be an increase in the magnitude of $ratio$. The deviation scoring algorithm is originally used in reconstruction signal, while we made the first improvement by applying such algorithm directly in wavelet packet coefficient. At the initial anomaly detection of time t, through the wavelet packet decomposition for the network traffic signals during $(t - HisWin, t)$, we get the wavelet packet coefficient on each scale, which we detect by means of deviation scoring algorithm, then we get to the conclusion that whether there is an anomaly as long as the value of ratio is beyond the alert threshold. At the confirmation of anomaly of time t, we reconstruct the wavelet packet coefficient on those scales and detect again by means of deviation scoring algorithm.

If we only use deviation scoring to detect anomalies, the low frequency anomalies which last for a long-time may be invalid. The main reason is that, under certain condition the low frequency anomalies will keep stable when rising to a certain magnitude, and when the detection window is far shorter than the duration of the low frequency anomalies, the deviation scoring will abruptly change in the both beginning and end stage of anomalies, as well it will remain in the middle stage. At this time, a long-time anomaly is taken as two short-time ones. To solve this problem, we define a mean scoring to detect this kind of anomaly. At time t, we compute the mean E_1 of the detection window $(t - DetWin, t)$, and the mean E_2 of the historical window $(t - HisWin, t)$. Set $ratio_E = E_1/E_2$ as the mean scoring, which represents the variation of the sample in the detection window compared to the historical normal data. Generally speaking, the mean scoring will be stably larger than 1 at the beginning of the low frequency anomalies. Hence, combing with this mean scoring, we can accurately decide the long-time low frequency anomalies. We will illustrate it in step 3 of simulation.

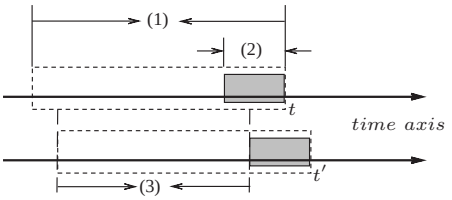


Fig. 3. The statistic detection algorithm based on sliding window.(1) HisWin, (2) DetWin, (3) The length of the redundancy data

B. Detection Model and The Double Thresholds Mechanism

The model of network traffic anomaly detection is shown in Figure 4. It can be divided into five parts in general, such as, wavelet packet analysis, initial anomaly detection, reconstruction of wavelet packet and confirmation of anomaly. The initial anomaly detection applies the double thresholds mechanism: they are two thresholds–alert threshold (T_a and T_{Ea}) and decomposition threshold (T_d and T_{Ed}), where $T_a > T_d, T_{Ea} > T_{Ed}$. If alert threshold is reached, we consider it as an anomaly, and if the decomposition threshold is reached, we consider it as could-be anomaly.

At first, we make multi-scale 1-level decomposition for the traffic signals and make wavelet packet decomposition for the coefficient of node [1,0] to the level 3. Now, we can detect the coefficient anomaly under different scales by means of statistic detection algorithm of the sliding window. When it is found that anomaly reaches the alert threshold in some certain scale of the former n levels, reconstruction detection starts immediately; if it is still anomalous, it alarms. When it is found that anomaly reaches the decomposition threshold in certain level of the n-th scale, decomposition continues to detect on the level $(n+1)$, decomposition will end until the anomaly reaches the alert threshold or the anomaly is below the decomposition threshold. The wavelet packet decomposition tree of our method is shown in Figure 5.

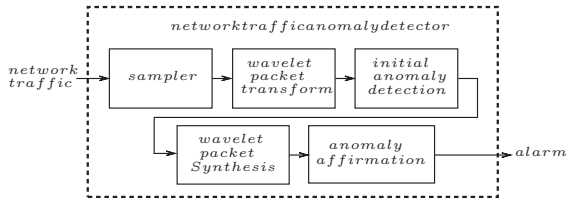


Fig. 4. The model of network traffic anomaly detection

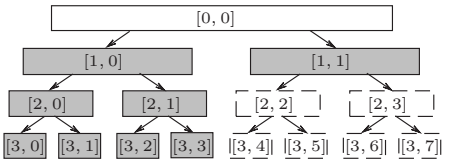


Fig. 5. The wavelet packet decomposition tree(level=3), and the fuscous panes are the coefficient of our method

C. Detection Steps

Functions of each module and the steps of detection are as follows, and the flow chart of detection algorithm is shown in Figure 6.

1. Generation of detection signal: We take packets gathered by router per unit time as the traffic signal, the sample interval is T_0 sec. If $f(n)$ is the value of the n-th sample, then:

$$f(n) = \begin{cases} 0, & n = 0. \\ (T_0 \cdot (n-1), T_0 \cdot n] \ packet \ numbers, & \\ otherwise. & \end{cases} \quad (16)$$

2. Wavelet packet analysis: Even the weak high-frequency anomaly could be detected from the node [1,1], so it is enough to make multi-scale decomposition to level 1, and then make wavelet packet decomposition from the node [1,0]. With the increase of the levels of the wavelet packet decomposition, the number of the wavelet packet coefficient halves. If the length of the detection series is N, the length of the wavelet packet coefficient series from level j would be $N/2^j$. Besides, since the wavelet packet decomposition is based on 2-abstract DWT, with the increase of the levels, the nodes of each level increase by way of 2^j. Therefore, the levels decomposed in the initial period are limited, and afterwards they will continue self-adaptive decomposition with the specific detection situations.

3. Initial anomaly detection: Make the initial detection on the wavelet packet coefficient of each scale, and check whether there is any anomaly at some moments on this scale, by the means of statistical detection algorithm. If $ratio > T_a$ or $ratio_E > T_{Ea}$(that is anomaly), go to step 4; if $ratio < T_a, ratio > T_d$ or $ratio_E < T_{Ea}, ratio_E > T_{Ed}$(that is could-be anomaly), we further decompose the wavelet packet coefficient and then detect them by step 3 again; if $ratio < T_d, ratio_E < T_{Ed}$, the could-be anomaly can be removed. Therefore, the decomposition levels are totally self-adaptive.

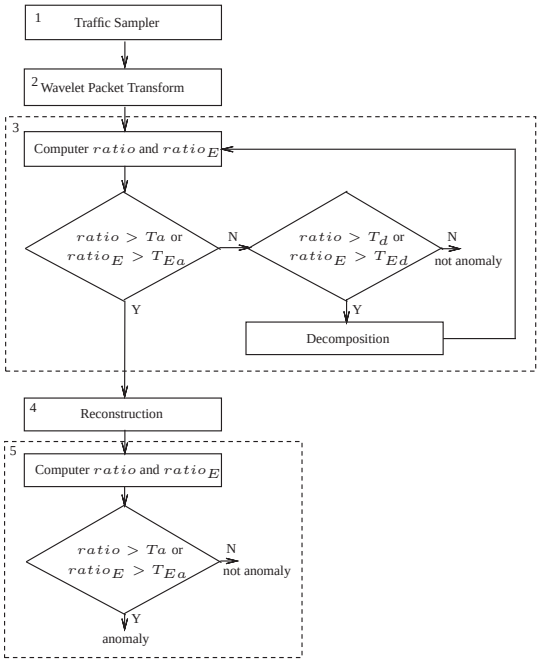


Fig. 6. The flow chart of the detection algorithm

4. Reconstruction of wavelet packet: The energy of anomalous signals mainly locates on those scales, where we detect the anomaly in step3. Therefore, we selectively reconstruct the wavelet packet coefficient on those scales. So in this way, the reconstructed signals can be distinct from original signals to greater degree.

5. Confirmation of anomaly: The anomaly we get from initial detection module may be a mis-detection, so we need to redetect the reconstructed signal to reduce the ratio of mis-detection. Furthermore, 2-abstract DWT has poor performance in time domain. We can get better performance in time location if we detect reconstructed signal. In our method, we set an alert threshold. If the detection result in reconstructed signal reaches the alert threshold, we regard it as an anomaly; otherwise, we recognize it as a mis-detection.

D. Detection Window

Generally speaking, the detection window is difficult to fix. if we choice a detection window at random, it dose not guarantee the detection effect. Therefore, we propose a method to fix the detection window by means of the center frequency of every wavelet domain.

In the initial anomaly detection, signals to be detected are the wavelet packet coefficient of each level. After making wavelet packet decomposition for traffic signals to the level j, there would be 2^j frequency band series with same bandwidth, the length of signals in each frequency band decrease to the $1/2^j$ of the traffic signals, the sampling interval is 2^j times of traffic signals. If the highest frequency of the traffic signals is f, the frequency range of the 2^j frequency band would be

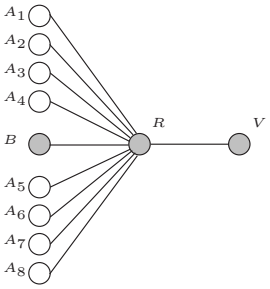


Fig. 7. The simulation topological

$$2^{-j}(i-1)f \sim 2^{-j}if, \tag{17}$$

Here, $i = 1, 2, \ldots, 2^j$ stands for the frequency band series of decomposition signals. We can approximately calculate the center frequency of each frequency band by

$$f_{cj}^{\ i} = 2^{-(j+1)}(2i-1)f, \tag{18}$$

If the sampling interval (for traffic signals)of the level 0 is Δ, that of the level j would be $2^j\Delta$. We take the data length corresponding to the 2^j times period as the size of the detection window by

$$DetWin_j^i = 2^{(j+1)}/[(2i-1)f \cdot \Delta], \tag{19}$$

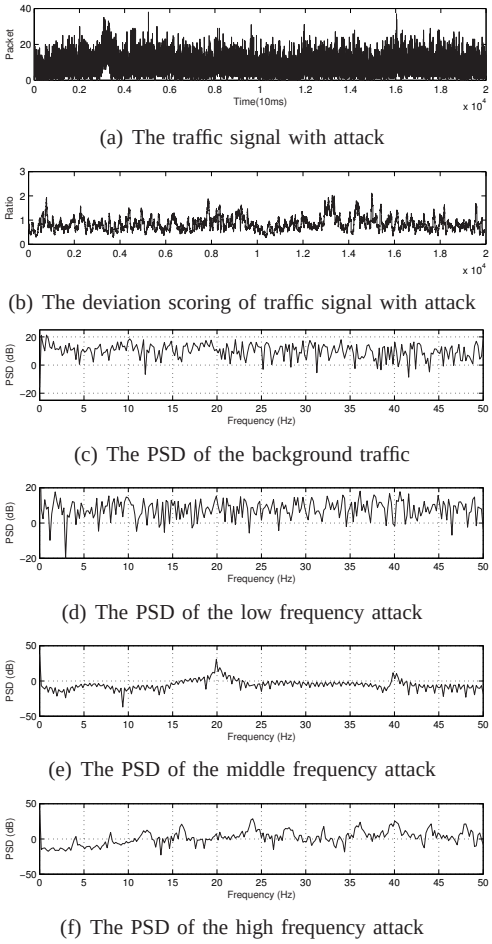
We can see from the above that the size of detection windows calculated on the basis of center frequency differs on all scales of at each level: it's little bigger under low-frequency detection and smaller under high-frequency detection. When anomaly is confirmed, the signals to be detected are reconstructed ones. To ensure that anomaly in all frequency ranges can be effectively detected, the detection window is set to be the maximum of the reconstructed detection window of wavelet packet coefficient.

In practice, the over small detection window would lead us to think the very short normal traffic increasing instantaneously as an anomaly, so in our simulation experiments, we fix the smallest detection window to be 6.

IV. Results

A. Simulation Backgrounds

In our simulation experiments, we adopt the data[19] as the background traffic, which were gathered by Lawrence Berkeley lab in University of California, Berkeley Institute; According to the principle of DDoS attack, we simulate 8 data source as attack source, which send a huge volume of traffic to a victim at the same time. The simulation topological is shown in Figure 7, in which B is the normal traffic data source, that is background traffic, $A_x(x = 1 \sim 8)$ is the data source of DDoS attack, R is the router before the Victim hostand V is the victim host. We use the DDoS attack as anomalous traffics, which was produced in NS2 simulation . And then we simulate that eight attack source host computers launch



(a) The traffic signal with attack

(b) The deviation scoring of traffic signal with attack

(c) The PSD of the background traffic

(d) The PSD of the low frequency attack

(e) The PSD of the middle frequency attack

(f) The PSD of the high frequency attack

Fig. 8. The traffic signal, detection result and PSD

attacks to the victim host computer during 100ms at random, if each attack source host computer sends data packets by exponential distribution with the average value of 10ms, it is a kind of low-frequency attack; if each attack source host computer sends 7,5,2,0,3 data packets respectively every 10ms , it is a sort of middle-frequency attack; if each attack source host computer sends 2,10,1,8,5 data packets respectively every 10ms, it is a kind of high-frequency attack. Attack 1 (starting time: 40-44s) added the background traffics is low-frequency attack; Attack 2 (starting time: 80-82s) is middle-frequency attack; Attack 3 (starting time: 160-161s) is high-frequency attack. The time interval of sampling is 10ms and the traffic data is shown in Figure 8(a). If we get deviational numeric (picking 30 from the detection window) in time domain traffics to detect, it cannot detect any attack at all, the result is shown in Figure 8(b), The choice of Wavelet packet filter is important in detection performance in the decomposition and reconstruction of traffic data. After weighing the effect of linearity, symmetry, vanishing moment, and localization, we select $Daubechies$(db6) filter. All the simulation experiments in this paper are operated on the computer with 2.4GHz Pentium 4 processor and 512M memory.

B. The Fast Wavelet Packet Algorithm

In order to prove the validity of the fast wavelet packet algorithm based on sliding window, we compute the wavelet packet coefficient of traffic signals by means of normal wavelet packet algorithm and the fast wavelet packet algorithm. We decompose the traffic signals to level 4, based on wavelet packet decomposition; the length of filter is 21 points, and the width of sliding window is fixed as 10000 points. As in the Figure 9, when the decomposition levels of wavelet packet and the width of sliding window are fixed, by using the fast wavelet packet transform algorithm, the time spent in computation of the wavelet packet coefficient increases with the time of smooth sliding, and when time interval approaches to the width of sliding window, the computation time of fast wavelet packet transform algorithm is close to that of normal wavelet packet algorithm, and by applying two algorithms, the wavelet packet coefficient are almost the same.

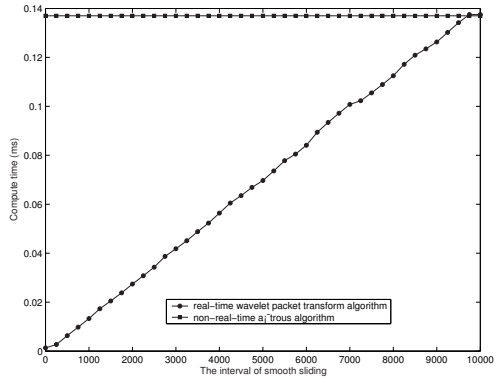


Fig. 9. The time spent in computation of the wavelet packet coefficient

C. The Network Traffic Anomaly Detection Algorithm

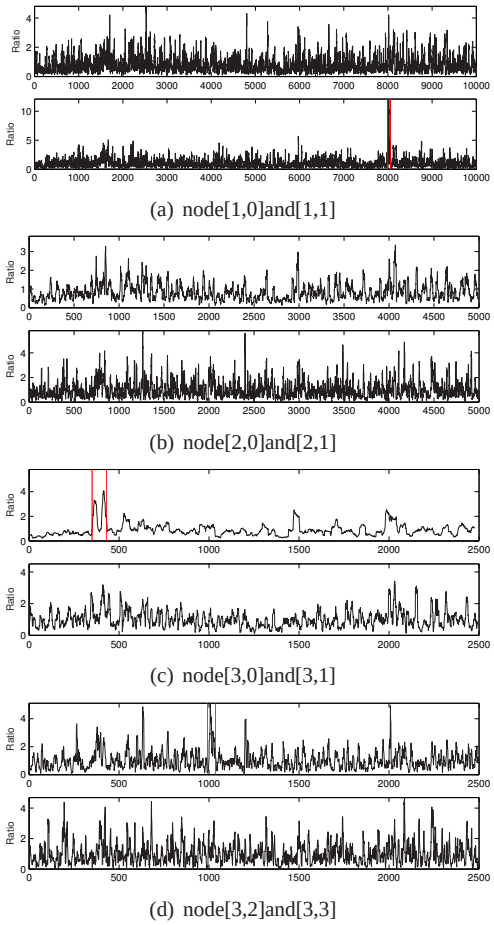
Through the research of historic traffics, make sure the threshold value in each scale before detection.

1.Make 1-level multi-scale decomposition to the traffic signals and initial anomaly detection on the wavelet packet coefficient in each corresponding level. As it is shown in Figure 10(a), Attack 3 surpasses the alert threshold at [1,1], so it is detected out. Then decompose node [1,0] to level 3; as it is shown in Figure 10(c) and 10(d), attack 1 and 2 surpass the alert threshold at [3,0] and [3,2] respectively, thus they are required to be further decomposed and detected. If we only use multi-scale decomposition, attack 2 hidden at [2,1],as is shown in Figure 10(b), can not be further detected, which will leads a mis-detection.

2. Decompose the coefficient [3,0],[3,2] of could-be attack into level 4, as in Figure 11. When finding attack 1 and 2 surpass the alert threshold at [4,0],[4,5], we consider there is an attack. Certain energy of attack 3 is concentrated at [4,4], so we should take it into consideration when making reconstruction.

3. Make reconstruction detection to ensure attack. Reconstruct [4,0] for attack 1, [4,5] for attack 2, [1,1] and [4,4] for attack 3, as the detection result shown in Figure 12(a),12(c) and 12(d), all of the three surpass the alert threshold. We find that only by using deviation scoring, we cannot detect out attack 1 completely. As is shown in Figure 11(a), a long-time anomaly is taken as two short-time ones. But combined with Mean Scoring, we can detect out that it is a durative attack, as in Figure 12(b).

Meanwhile it can prove that our choosing method of detection window can well detect the attacks of various frequencies, for instance, in Figure 12, the detection window of [4,0] and [4,5] is 64 and 6,but if we fix the detection window at value30, the result is shown in Figure 13, which is not so good as in Figure 12.



(a) node[1,0]and[1,1]

(b) node[2,0]and[2,1]

(c) node[3,0]and[3,1]

(d) node[3,2]and[3,3]

Fig. 10. The deviation scoring of 1,2,3-level wavelet packet coefficient

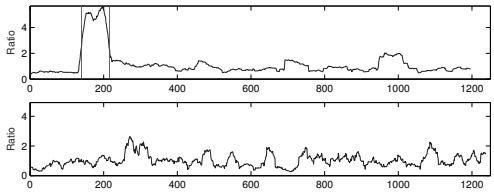
V. Conclusion

In this paper, we propose a new mechanism of network traffic anomaly detection based on wavelet packet analysis. It can select different time-frequency resolution to decompose adaptively according to the characteristics of traffic signal. Using our method, we can locate time-frequency domains and get the faint signal effectively. According to simulation results, this mechanism is proved to be feasible, and it possesses some merits as follows: (1). It can effectively detect the long-time durative anomalous traffic and the short-time sudden-

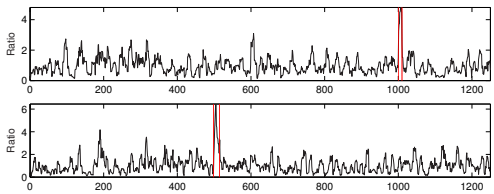
changing one, and also it can effectively detect middle-high frequency attack traffic which can not be checked out by the network traffic anomaly detection based on multi-resolution analysis. By means of decomposing threshold, it has avoided the blindness of wavelet packet decomposition, and solved the problem of decomposing scale's self-adaptive selection. (2). by means of adaptive reconstructing the wavelet packet coefficient in different wavelet domains which include anomaly, our method is able to confirm the characteristics of anomaly and enhance the reliability of detection. (3). Applying the fast wavelet transform algorithm to the network traffic anomaly detection has greatly reduced the computation complexity.(4). Using the center frequency of different frequency bands to compute the size of the corresponding detection windows, it has solved the problem of adaptive selection of detection windows.

ACKNOWLEDGMENTS

The work described in this paper was supported by a grant from NSFC (Project No. 60572092)



(a) node[4,0]and[4,1]



(b) node[4,4]and[4,5]

Fig. 11. The deviation scoring of level 4 wavelet packet coefficient

REFERENCES

[1] L. Lewis and G. Dreo, "Extending trouble ticket systems to fault diagnosis," *IEEE Network*, vol. 7, pp. 44–51, 1993.

[2] L. Lewis, "A case based reasoning approach to the management of faults in communication networks," *Proc.IEEE INFOCOM*, vol. 3, pp. 1422–1429, 1993.

[3] I. Katzela and M. Schwarz, "Schemes for fault identification in communication networks," *IEEE/ACM Trans. Networking*, vol. 3, pp. 753–764, 1995.

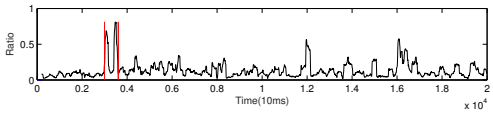
[4] I. Rouvellou and G. Hart, "Automatic alarm correlation for faultidentification," *Proc. IEEE INFOCOM*, pp. 553–561, 1995.

[5] F. Feather and R. Maxion, "Fault detection in an ethernet network using anomaly signature matching," vol. 23, pp. 279–288, 1993.

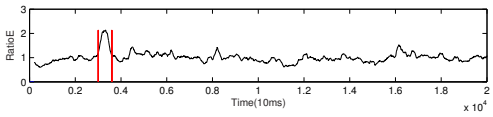
[6] S. Papavassiliou, M. Pace, and L. Ho, "Implementing enhanced network maintenance for transaction access services: Tools and applications," vol. 1, 2000.

[7] M. Thottan and J. Chuanyi, "Anomaly detection in ip networks," *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, vol. 51, no. 8, 2003.

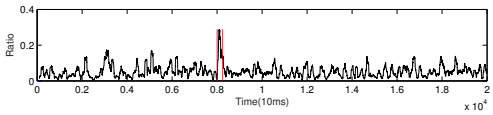
[8] C.-M. Cheng, H. Kung, and K.-S. Tan, "Use of spectral analysis in defense against dos attacks," *Proceedings of IEEE GLOBECOM 2002*, 2002.



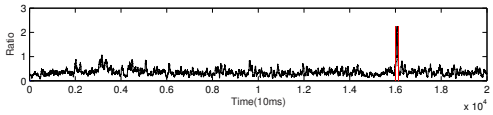
(a) The deviation scoring of attack 1



(b) The mean scoring of attack 1

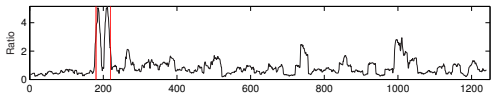


(c) The deviation scoring of attack 2

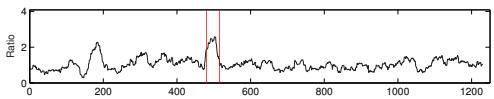


(d) The deviation scoring of attack 3

Fig. 12. The deviation scoring and Mean Scoring of the reconstructed signal



(a) node[4,0]



(b) node[4,5]

Fig. 13. The deviation scoring of level 4 wavelet packet coefficient

[9] W. Allen and G. Marin, "On the self-similarity of synthetic traffic for the evaluation of intrusion detection systems," *Proceedings of the 2003 Symposium on Applications and the Internet (SAINT03)*, 2003.

[10] A. Lakhina, M. Crovella, and C. Diot, "Diagnosing network-wide traffic anomalies," *ACM SIGCOMM*, 2004.

[11] A. Lakhina, K. Papagiannaki, M. Crovella, C. Diot, E. Kolaczyk, and N. Taft, "Structural analysis of network traffic flows," *ACM SIGMETRICS*, 2004.

[12] A. Lakhina, M. Crovella, and C. Diot, "Characterization of network-wide anomaliesin traffic flows," BUCS-2004-020, Boston University, Tech. Rep., 2004.

[13] B. BR, "Multi-scale analysis and modeling using wavelets," *Journal of Chemometrics*, vol. 13, 1999.

[14] V. Alarcon-Aquino and A. Barria, "Anomaly detection in communication networks using wavelets," *IEEE Proc-Commun*, vol. 148, no. 6, 2001.

[15] A. Ramanathan, "Wades: A tool for distributed denial of service attack detection," *TAMU-ECE-2002-02, Master of Science Thesis*, 2002.

[16] P. Barford, J. Kline, D. Plonka, and A. Ron, "A signal analysis of network traffic anomalies," *InternetMeasurement Workshop*, 2002.

[17] S. S. Kim and A. Reddy, "Detecting traffic anomalies at the source through aggregate analysis of packet header data," *Proceedings of Networking*, 2004.

[18] L. Lan and L. Gyungho, "Ddos attack detection and wavelets," *Telecommunication Systems*, pp. 435–451, 2005.

[19] *Internet Traffic Archive*, http://ita.ee.lbl.gov/index.html.