Network Anomaly Detection based on Traffic Prediction

Fengyu Wang, Bin Gong, Yi Hu, Ningbo Zhang
College of Computer Science and Technology
Shandong University
Jinan, P. R. China
e-mail: {wangfengyu, gb, huyi, zhangnb}@sdu.edu.cn

Abstract—**As the development of Internet, it is more and more difficult to detect anomaly promptly and precisely. In this paper, we proposed an anomaly detection algorithm based on the predicting of multi-level wavelet detail signals synchronously. Firstly, process the time series of traffic with non-decimated Haar wavelet transform and produce detail signals. Secondly, predict the detail signals of wavelet transform and get the residual ratio series, which can expose the anomaly more obviously than original signal. Finally, based on principal of "3σ" of normal distribution, abrupt changes can be detected. Along with the arriving of traffic data, this algorithm detects anomaly on several time-scales recursively without delay. So this algorithm can detect traffic anomaly more precisely and promptly. Analysis and experiments reveal that this algorithm can detect anomalies effectively.**

Keywords **—** ***anomaly detection; network traffic prediction; wavelet transform***

I. INTRODUCTION

With the rapid development of Internet network, large-scale security events happen frequently. On the one hand, the security related events (i.e. the spread of Internet worms, DDoS attacks) consume large amount of Internet bandwidth and CPU resources, which may destroy hosts, slow down Internet performance and even collapse the whole Internet. On the other hand, the failure of network and flash crowd of connections will also make flows of Internet change rapidly, which affects the normal services on the Internet. The anomaly of network can be characterized as abrupt change of traffic, so monitoring traffic on the Internet is an effective method to guarantee Internet performance. Anomaly detection of network traffic is to construct normal pattern of network traffic and alarm if the diversity between the parameter of real time traffic and the pattern constructed previously exceeds pre-defined threshold [1].

The procedure of Traffic anomaly detection can roughly be separated into two steps: traffic pattern modeling and anomaly detection. Traffic pattern modeling is to extract network traffic profiles at different levels and anomaly detection is to determine whether the network traffic is natural by comparing measured traffic parameters with the threshold produced with traffic model. The anomaly of network traffic could be presented more obviously through choosing appropriate traffic patterns from different angles (e.g., some detection methods observe specified fields of network protocols in order to detect anomalies more accurately). However, it needs many resources to detect all potentially anomalies for the varieties of intrusions.

The basic idea of anomaly detection according to the specified details of network traffic is dividing the whole traffic into separate levels in order to expose the abnormal traffics. The wavelet transformation can split the observation objects in time domain into different levels in spectral domain, which can separate abnormal traffic and normal traffic as well as avoiding the influence of measurement variations. However, the computing complexity of the current detection algorithms based on wavelet transformation is relative high, which make it difficult to implement real-time detections. In this paper, we decompose traffic time series with non-decimated Haar wavelet transform recursively and fit the detail signals at each time-scale level with AAR (adaptive auto-regressive) model based on RLS (recursive least square). The distribution of residual series produced by AAR model is Gaussian white noise, on which the traffic abrupt changes can be accurately detected. This algorithm can meet the requirements on accuracy and real-time of anomaly detection in high-speed network environments.

The rest of this paper is organized as follows. In Section II, the related research works are presented, and in Section III, we describe our anomaly detection algorithm based on multi-scale traffic prediction and analyze its performance. In Section IV, we evaluate the accuracy and feasibility of our methods with collected network traces. In Section V, the effect of algorithm is verified with a real DDoS attack dataset. Finally, the work is concluded in Section VI.

II. RELATED WORKS

Many researches have been developed on detecting anomaly of network traffic. The main methods used in these algorithms can be roughly classified into the following types: statistic, machine learning and data mining etc. Many of these algorithms focus on detecting anomalies on LAN or Intranet and consume lots of computing resources to achieve higher accuracy, such as machine learning based detection and data mining based detection. The computing efficiency of these algorithms can not meet the requirements of on-line anomaly detection in backbone network links, which transfer much more traffic volume in higher speed.

This work was supported in part by the National Science Foundation of China (60803142).

978-0-7695-3825-9/09 $26.00 © 2009 IEEE
DOI 10.1109/EmbeddedCom-ScalCom.2009.86

Statistic-based anomaly detection approaches hold some advantages, such as high efficiency and wide applicability, so they have been widely adopted in high-speed network anomaly detection. According to the way of measuring the traffic profile, we can further classify the anomaly detection approaches into two categories: (1) Analyzing specific detail characteristics of the network traffic. When an attack is launched, obvious changes would be observed in the value of some specified protocol fields or the ratio of them, so this kind of approaches can detect some specific attacks exactly. In [2], Wang et al. take the difference in the number of SYNs and FINs collected within one sampling period as time series data and use CUSUM method to detect SYN flooding DDoS attack. Kompella detected the imbalance through a partial completion filter, which can detect claim-and-hold attacks in high-speed network [3]. However, if there are n potential metrics related to various attack strategies, then at most, there are 2^n subsets should be detected [1], the computing complexity would be unacceptable. (2) Analyzing the time series of packet count or flow count in one fixed measurement interval. This kind of approaches detect anomaly with time series models or statistical principles after some appropriate pre-processing. In [4], Throttan and Ji predict the network traffic with AR model and compute the GLR (Generalized Likelihood Ratio) of residuals in two sliding windows to detect anomaly. ZOU et al. [4] detect anomalies with ratio of residuals based on the AR model. In [6], the AAR model is used to detect the abrupt changes on FCD (Flow Connection Density). This kind of approaches is not constrained to some specific traffic anomalies, but due to the dynamics and complexity of high-speed network traffic, it's difficult for these models to characterize the network traffic.

It is shown that LAN/WAN packet arrival processes could be modeled more accurately using self-similar processes [7]. Wavelet transform based network traffic model can analyze signal on various scales by rescaling and translation, which could describe both long-term dependences and short-term dependences in time domain, so it could characterize network traffic more accurately [8]. In [9] and [10], network traffic time series are pre-processed with wavelet decomposition and detected on various time scales, which improves the accuracy of network traffic anomaly detection. However, the information after time point t is needed to compute the wavelet coefficients at t in the wavelet transform, so the anomaly detection is postponed in a way on various time scales. Furthermore, the sliding window used in these detections can delay the anomaly alarming.

III. Anomaly Detection Based on Prediction

There are two ways to optimize traffic anomaly detection of high-speed network. One is to find a traffic profile factor which can expose the abnormal change more clearly from some aspects; the other is to found a more accurate and efficient traffic model. From the second way, we propose an anomaly detecting algorithm based on multi-scale prediction. Firstly, pre-process the time series of traffic with non-decimated Haar wavelet transform algorithm, then fit various detail signal parts with AAR model and produce residual series, which amplify the abrupt changes and the anomaly is detected on the basis of residual ratio.

A. Pre-process with Wavelet Transform

Pre-processing the traffic series with wavelet transform can improve the traffic model in three aspects. Firstly, the network traffic is self-similar and the wavelet transform can eliminate the correlation, so the short-dependence model can fit the decomposed network traffic time series more accurately. Secondly, the original signal is decomposed into several detail signals and an approximate signal, which can isolate the abrupt changes from the ambient parts. Finally, the attacks that increase their traffic slowly can be revealed through detecting anomaly in various time scales simultaneously.

Traditionally, pyramid algorithm is used in wavelet decomposition and reconstruction. The number of points is half of the prior level as decomposing one time in pyramid algorithm, so we cannot simply relate information at a given time point at the different scales. With somewhat greater difficulty, however, this goal is possible. What is not possible is to have shift invariance. This means that if we had deleted the first few values of our input time series, then the output of wavelet transform would not be the same as heretofore. So we adopt the à trous wavelet transform [11].

The à trous wavelet transform is a redundant transform. A redundant transform based on an n-length input time series, then, has an n-length resolution scale for each of the resolution levels. It is easy to relate information at each resolution scale for the same time point. More importantly, we have shift invariance. Given a time series $x(t)$, the scale coefficients can be produced with discrete low pass filter:

$$c_0(t) = x(t), \quad c_{j+1}(t) = \sum_{l=-\infty}^{\infty} h(l) c_j(t + 2^j l) \tag{1}$$

and the wavelet coefficients can be computed with scale coefficients:

$$d_{j+1}(t) = c_j(t) - c_{j+1}(t) \tag{2}$$

So the result of à trous wavelet transform is a set $\{d_1, d_2, \ldots, d_J, c_J\}$, where d_j (j=1,2,…,J) are the detail signals of various time scales and c_J is approximate signal. The original signal can be reconstructed with detail signals and approximate signal :

$$x(t) = c_J(t) + \sum_{j=1}^{J} d_j(t) \tag{3}$$

The non-decimated Haar algorithm uses the simple filter h=(1/2,1/2). The scale coefficients and wavelet coefficients can be computed according to formula (1) and (2):

$$c_0(t) = x(t) \tag{4}$$

$$c_{j+1}(t) = 0.5(c_j(t - 2^j) + c_j(t)) \tag{5}$$

$$d_{j+1}(t) = c_j(t) - c_{j+1}(t) \tag{6}$$

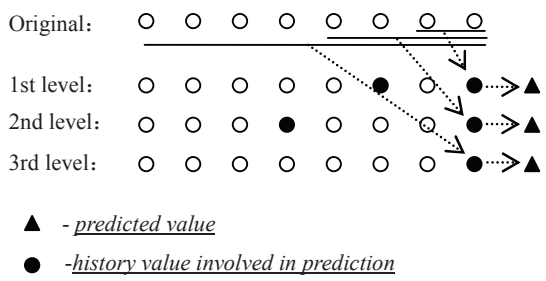


Fig. 1 Wavelet transform and detail signal prediction

Along with the arriving of new data, it is not necessary to re-compute prior coefficients, so this wavelet transform is adaptive to online computing. When new data arriving, the data points related with decomposing in various levels are labeled with lines in Figure 1. We contain the history data in a sliding window, including original signal, detail signals and approximate signals. The size of sliding window is related to the resolution level of wavelet transform and the order number of AAR model.

B. Predicting Detail Signals with AAR Model

The original traffic series is decomposed to several detail signals and an approximate signal by wavelet transform. Because the detail signals are not stationary process in some time scales, the parameters should be updated sometimes if we use AR model to fit the detail signals. AAR model [12] can address this issue. An AAR model with order *p* is written as

$$\begin{aligned} y(t) &= a_1(t)*y(t-1)+...+a_p(t)*y(t-p)+\varepsilon(t) \\ &= a(t)^T*Y(t-1)+\varepsilon(t) \end{aligned} \tag{7}$$

where $\varepsilon(t)$ is Gaussian white noise, $a_i(t)$ (i=1,2,…,p) is real numbers, y(*t*) is random variable and *t* is time or sequence number of sample. The difference with (stationary) autoregressive (AR) model being, that the AAR parameters vary with time.

Suppose that the parameter vector of AR model at *t* time point is

$$a(t) = [a_1(t),\ a_2(t),\ \ldots,\ a_p(t)] \tag{8}$$

The corresponding sample vector is

$$Y(t-1) = [y(t-1), y(t-2), ..., y(t-p)] \tag{9}$$

and the estimated value of a(t-1) is $\hat{a}(t\text{-}1)$, then the error of one step prediction is

$$e(t)=y(t)-\hat{a}(t\text{-}1)^T * Y(t-1) \tag{10}$$

The estimation of AAR model parameters based on RLS is

$$Q(t) = Y(t-1)^T * A(t-1) * Y(t-1) + 1/(1-UC) \tag{11}$$

$$k(t) = A(t-1) * Y(t-1) / Q(t) \tag{12}$$

$$\hat{a}(t) = \hat{a}(t\text{-}1) + k(t)^T * e(t) \tag{13}$$

$$\begin{aligned} A(t) = A(t-1) - k(t) * Y(t-1)^T * A(t-1) \\ + UC * A(t-1) \end{aligned} \tag{14}$$

where UC is the update coefficient, k(*t*) is the update gain vector, A(*t*) is the correlation matrix of samples with order *p*. Given initial values of UC, A(t-1) and $\hat{a}(t\text{-}1)$, the model parameters can be estimated according to formula (11)-(14). The correlation matrix and the update gain vector barely affecting the results, so they can be set to zero vector and identity matrix. The update coefficient can affect the convergence speed of parameter estimation greatly. Along with the decreasing of update coefficient, the vibration of will lessen and the delay of parameter identification will increase. So the value of update coefficient should be set according to the experiment results. We use UC=0.02 in this paper.

There are some redundant data in the detail signals of non-decimated Haar wavelet transform, so we can predict the value of N+1 time point with

$$d_{j,N-2^j(k-1)}$$

in various time scale *j*. To decrease the computing cost, we select the AAR model order *p*=2. In Figure 1, the circles filled with black are history data used in AAR prediction and the triangles is the prediction value.

C. Anomaly Detection

Based on the analysis in Section B, we know that the non-redundant detail signals can be approximated with Gaussian White Noise. These detail signals are stationary series with mean 0. When abrupt change happened, the residual would deviate from the statistical law. According formula (10), we can get the residual series e(t). Suppose the mean and variance of e(t) is $\mu(=0)$ and σ^2 respectively. Then the probability distribution of e(t) obeys to:

$$P\{x_1 < e(t) < x_2\} = \Phi\left(\frac{x_2-\mu}{\sigma}\right) - \Phi\left(\frac{x_1-\mu}{\sigma}\right) \tag{15}$$

The $\Phi(x)$ is the distribution function of standard normal distribution. We can deduce the factual distribution principle of residuals:

$$P\{-3\sigma < e(t) < 3\sigma\} = 99.74\% \tag{16}$$

This is the "3σ" principle of normal distribution, which means more than 99% of the probability mass is within +/-λ σ (λ=3)from the mean. According this law, we can detect the anomaly based on the history residuals preserved in a sliding window. The threshold of λ can be modulated according to the various situations. Its value ranges from 3 to 4.

When an anomaly happened, it would affect the consequent anomaly detection. To judge anomaly based on normal data, we replace the abnormal wavelet coefficient with the mean (i.e. 0) and remove the residual from sliding window when anomaly happened. Simultaneously, we construct another sliding window to save these abnormal data including wavelet transform series and residual series. When the length of abnormal data is enough to be used to detect anomaly, we use this set of sliding windows to detect

anomaly alternatively. This scheme can keep the validity of anomaly detecting when a traffic change has continued for a long time.

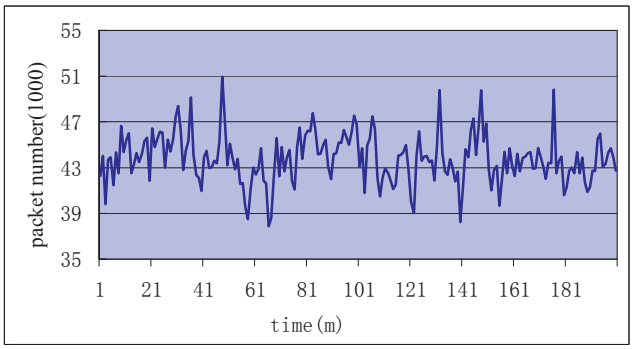
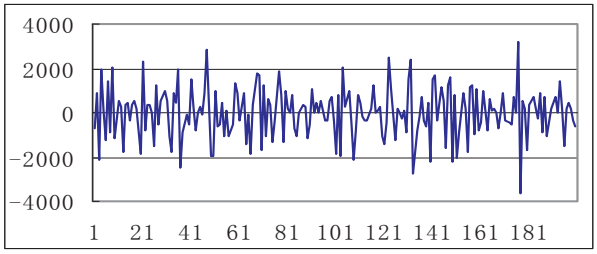
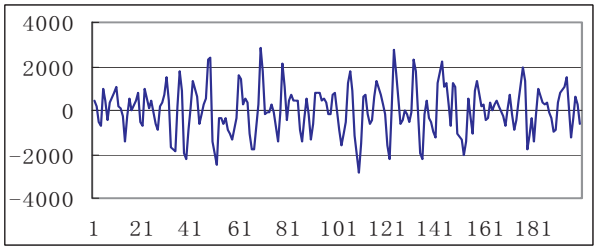


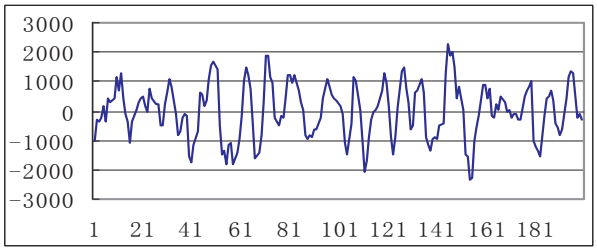
Fig. 2 Traffic trace of Auckland



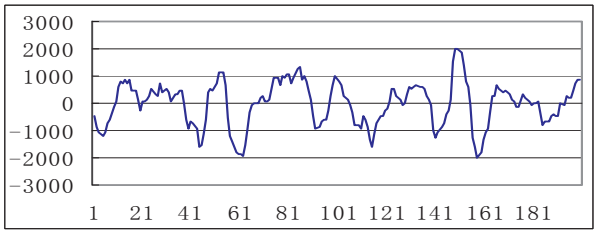
a. first level



b. second level



c. third level



d. fourth level

Fig. 3 Detail signals of wavelet decompose (4 levels)

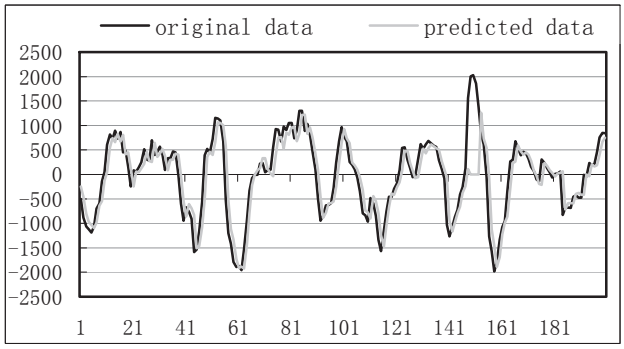


Fig. 4 Prediction of detail signal (the fourth level)

IV. Analysis Based on Real Network Trace

We use network traffic trace from the access point to the Internet of Auckland University on an OC-3c link in this Section. It is measured by NLANR [14], whose traffic traces are adopted by many researchers. Figure 2 shows the traffic time series, the measurement interval is one minute.

Network traffic time series is self-similar in some time scales. The detail signals of self-similar signal are stationary signal in broad sense and its mean is zero. We produce these detail signals with non-decimated Haar wavelet transform and show the first four levels in Figure 3. We check the stationary of them with run test algorithm. Along with the increasing of signal level, the stationary coefficient decreased. So the time-variant AR model is adaptive to the detail signals. We use AAR model based on RLS [12].

After predicting the detail signals with AAR model, we can get the residual series, which are Gaussian white noises. Figure 4 shows the forth level detail signal and its one step prediction. Figure 5 shows the forth level detail signal and its residual series. A sharp change point is labeled with a circle. It is indicated that the abrupt change can be more obvious in residual series than in detail signals. This is because the

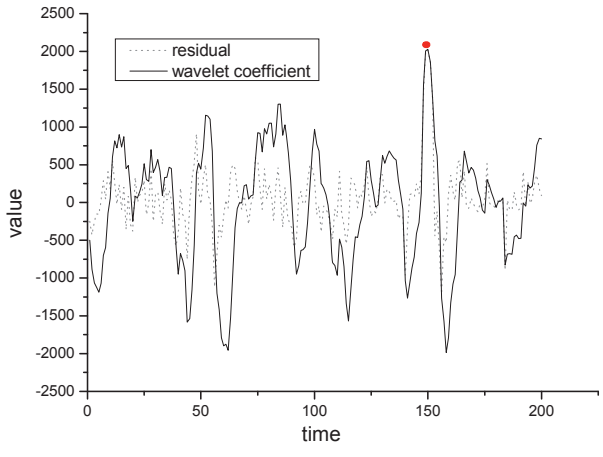


Fig. 5 Residual series (the fourth level)

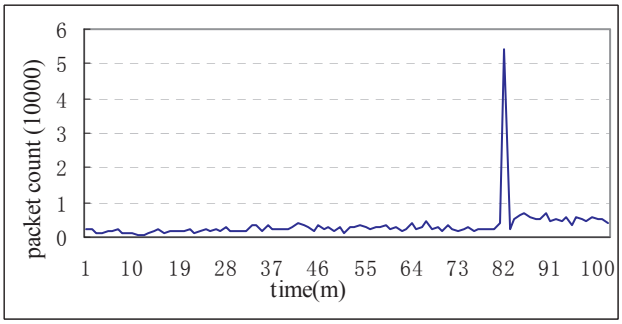


Fig. 6 DDoS traffic trace series

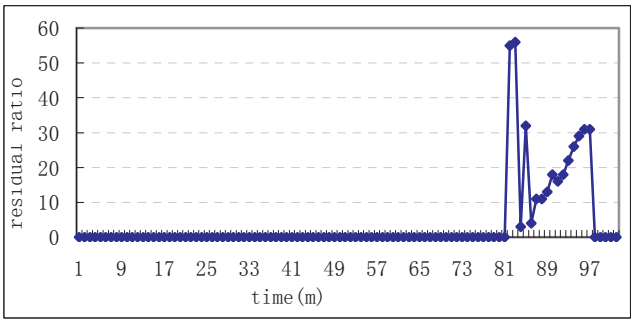


Fig. 7 Residual ratio series of traffic trace

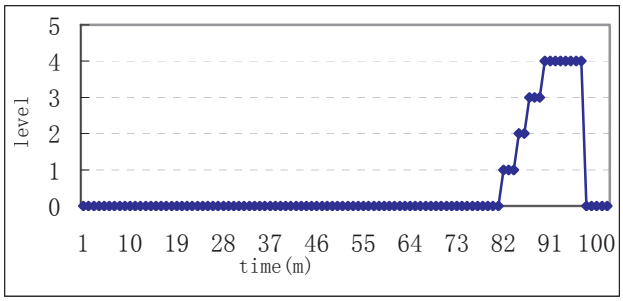


Fig. 8 Detail signal level where anomaly detected

normal variations can be predicted more accurately than abrupt changes, so the gap between the residual of them is enlarged.

V. Anomaly Detection Experiment

We adopt LLDOS2.0.2 [15] of MIT as the DDoS dataset. Figure 6 shows the packets number of one minute in LLDOS2.0.2. The residual ratio is shown in Figure 7, where the abrupt change of residual ratio can be easily detected on detail signals of wavelet transform. Figure 8 is the first one of wavelet detail signals where the anomaly is detected. The experiment indicates that the time point of detecting anomaly is distinct in different time scales, so our algorithms can detect the anomaly of network traffic more promptly. However, the local abnormal signals are reflected in several time scales, which make it hard to determine the end time of attacks. Whatever, it is more important to discover the attacks accurately in time.

VI. Conclusions

Anomaly detection of network traffic is always the open problems in network managements. It needs effective traffic models as well as balances between false positive and false negative. Although the accuracy of anomaly detection is still not satisfying, its importance in detecting unknown intrusions is not substitutable.

Along with the developing of high-speed network, new challenges are proposed for the efficiency of anomaly detection algorithm. Most of current detection methods are not suitable for high-speed network environments. In this paper, we combine non-decimated Haar wavelet recursive transform, ARR model based on RLS, and residual ratio together to construct a new anomaly detection algorithm. The experiments indicate that this algorithm can split normal traffic changes into several signal levels so as to expose the abrupt changes of traffic more obviously, which can improve the accuracy of detection effectively. Our algorithm can meet the needs of anomaly detection in high-speed network well both in processing speed and accuracy.

References

[1] QING Si-han, JIANG Jian-chun, MA Heng-tai et al. Research on intrusion detection techniques: a survey. Journal of China Institute of Communications, 2004,(7):19-29.

[2] Wang, H., Zhang, D., and Shin, K. G. Detecting syn flooding attacks. In Proceedings of IEEE INFOCOM (2002).

[3] R. R. Kompella, P. Singh, and P. Varghese, On scalable attack detection in the network. IEEE/ACM Transactions on Networking, 2007,15(1):14-25 .

[4] Throttan M, Ji C. Adaptive thresholding for proactive network problem detection. IEEE International Workshop on Systems Management. Newport , Rhode Island , 1998. 108-116

[5] ZOU Bo-xian. A Real Time Detection Method for Network Traffic Anomalies, Chinese Journal of Computer, 2003, 26(8): 940-947.

[6] Sun Qin-dong, Zhang De-yun, Gao Peng. Detecting Distributed Denial of Service Attacks Based on Time Series Analysis. Chinese Journal of Computer, 2005.28(5):767-773.

[7] Vern Paxson, Sally Floyd. Wide area traffic- the failure of Poisson modeling. IEEE/ACM Transactions on Networking, 1995, 3, 226-244.

[8] G. Wornell, Wavelet-based representations for the 1/f family of fractal process, Proc. IEEE, 1993, 81: 1428-1450.

[9] Alarcon V, Barria J A. Anomaly detection in communication networks using wavelets. IEE Proceedings Communications, 2001, 148(6):355-362.

[10] Barford P., Kline J., Plonka, D., and Ron, A. A signal analysis of network traffic anomalies. In Proceedings of ACM SIGCOMM Internet Measurement Workshop (2002).

[11] G. Zheng, JL Starck, J. Campbell and F. Murtagh, The wavelet transform for filtering financial data streams, Journal of Computational Intelligence in Finance, 1999, 7(3): 18-35.

[12] Schlogl A., Robert S. J . , Furt scheller G. P. A criterion for adaptive autoregressive models. In Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society , Chicago , 2000 , 1581-1582.

[13] Brutlag, J. D. Aberrant behavior detection in time series for network service monitoring. In Proceeding of the 14th Systems Administration Conference (2000)，pp. 139–146.

[14] NLANR Trace Archive. http://pma.nlanr.net/Special/, 2008.

[15] LLDOS Dataset. http://www.ll.mit.edu/IST/ideval/index.html, 2008.