The New Detection Algorithms for Network Traffic Anomalies

Jun Lv[2] Xing Li[1] Tong Li[2]
[1]China Education and Research Network, Tsinghua University,
Beijing 100084, China
[2]The Academy of Armoured Forced Engineering, Beijing, China

*Abstract–***Network traffic anomaly detection is a difficult problem in network management. This paper presents a Wavelet Generalized Likelihood Ratio (WGLR) algorithm and an Error Performance Detection (EPD) algorithm to solve this problem. WGLR algorithm combines Generalized Likelihood Ratio (GLR) algorithm and wavelet transform method, and captures the failure point in real time. Error performance Detection (EPD) algorithm is based on the prediction error of the traffic model and regards the error as the statistical variable, comparing with the threshold, which will detect the anomalous change in the signal without test window delay. Simulation and network traffic experiment has demonstrated that the algorithm has better performance in fault detection.**

1 Introduction

Internet is a large-scale complicated system. Due to the dynamic characteristic, it is difficult to detect the network abnormality and predict the time when the fault will happen. Therefore, many scholars have investigated the problems and suggested some methods to solve it.

- We usually preset a threshold by historical data, and compare the current data to this threshold. If the current data exceeds the threshold, the alarm will be generated, as described in Maxion[1].
- GLR (Generalized Likelihood Ratio) is commonly used in network problem detection [3]. This method considers three time window R(t), S(t), C(t), and using AR (Autoregressive) model to calculate the joint likelihood ratio of residual error in each time window. The edge point which exceeds the threshold is regarded as anomaly point.
- Jun Jiang[2] proposed a prediction method to detect the network server performance.

In this paper, we propose a new algorithm to detect the network anomaly. It combines the Generalized Likelihood Ratio (GLR) algorithm and wavelet transform method, and detect the fault in real time. By monitoring the port traffic and execute the new algorithm, we could be able to find and even predict some port traffic anomaly which can not be detected by threshold method, as well can not be detected from the overall network traffic.

To verify the effective of the algorithms, we simulate the algorithm and implement it on CERNET network. Simulation and the experimental results show that the new algorithm has an advantage of high accuracy. It determines the failure point to a nicety. Simplicity and reliability of the algorithm make it an attractive approach for online implementation.

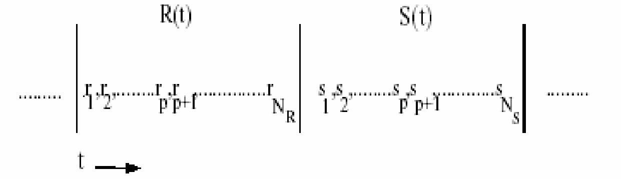
CERNET is a large network. We monitor the backbone traffic, collecting the original data, which is transmitted between CERNET and Beijing Internet Exchange Centre. Another traffic data comes from netflow, which has been collected at the CERNET international gateway, and obtained from CISCO router.

This paper has been arranged as follows: Section 2 discusses anomaly detection algorithm. In Section 3, Section 4 we will simulate and implementation the algorithm, comparing the performance of different methods. Conclusion and results will be presented in Section 5. We will give a summary in Section 6.

2 Anomaly Detection Algorithm

2.1 GLR Algorithm

Considering three sliding time windows: R(t), adjacent window S(t), and joint window L(t) (R(t) and S(t)), modelling the observations with AR(Autoregressive) model in three windows, calculating the residual error and decide a statistical variable for network failure detection[3].


R(t)
S(t)
$r_1\, r_2 \ldots r_p\, r_{p+1} \ldots r_{N_R}$
$s_1\, s_2 \ldots s_p\, s_{p+1} \ldots s_{N_S}$
t

Calculating likelihood ratio λ :

$$-\ln\lambda = N_R{}'(\ln\hat{\sigma}_p - \ln\hat{\sigma}_R) + N_S{}'(\ln\hat{\sigma}_p - \ln\hat{\sigma}_S) \tag{1}$$

threshold h : $-\ln\lambda > h \Rightarrow H_1$

$$-\ln\lambda \le h \Rightarrow H_0$$

Segment boundaries where the threshold was exceeded were considered to be change points.

Once a change is detected a second hypothesis test was conducted to determine whether a specific change point detected by the first hypothesis test.

The two hypotheses are H_0 with a distribution of $N(0,\sigma_0{}^2)$ implying that no change is observed, and H_1, with a distribution of $N(0,\sigma_R{}^2)$ implying that a change is observed. Set the threshold η to detect the fault.

GLR algorithm has some shortage as follows:

(1) Too much test window, complicated computation.

(2) Anomaly detection is not real time, has a delay of the test window, which will affect the accuracy of the detection.

So we present the wavelet GLR algorithm (WGLR).

Proceedings of the Sixth International Conference on Networking (ICN'07)
0-7695-2805-8/07 $20.00 © 2007 IEEE

2.2 Wavelet GLR Algorithm (WGLR)

This algorithm can be separated into two steps: the first step is calculating the likelihood ratio and detecting the abnormal; the second step is to diagnose the failure point.

2.2.1 Calculate the likelihood ratio

The first stage is finding the abnormal occurred. Considering only one time window R(t),

$$R(t) = \{r_1(t), r_2(t), ...r_{N_R}(t)\}$$

$\widetilde{r}_i(t) = r_i(t) - \mu$, μ is the mean of the segment $R(t)$

Now $\widetilde{r}_i(t)$ can be modelled as an AR process (p=2).

$$\varepsilon_i(t) = \sum_{k=0}^{p} \alpha_k \widetilde{r}_i(t-k)$$

where $\alpha_k = \{\alpha_1, \alpha_2, ...\alpha_p\}$ are the AR parameters and $\varepsilon_i(t)$ is assumed to be white noise. The joint likelihood of the residual time series was obtained as

$$p(\varepsilon_{p+1},..,\varepsilon_{N_R} / \alpha_1,..\alpha_p) = \left(\frac{1}{\sqrt{2\pi\sigma_R^{2}}}\right)^{N'_R} \exp\left(\frac{-N'_R \hat{\sigma}_R^{2}}{2\sigma_R^{2}}\right) \quad (2)$$

From Equation (2) and assuming that each sample of the residual error $\varepsilon_i(t)$ is drawn from an N(0, σ_R^{2}) distribution[7], where σ_R^{2} is the variance of the residual in segment R(t), and $N'_R = N_R - p$ and $\hat{\sigma}_R^{2}$ is the covariance estimate of σ_R^{2}.

The two hypotheses are H_0 with a distribution of N(0, σ_0^{2}) implying that no change is observed, and H_1, with a distribution of N(0, σ_R^{2}) implying that a change is observed.

The normal mean and variance were computed from the normal data (over a twenty four hours)

For H_0 :

$$p(\varepsilon_{p+1},..,\varepsilon_{N_0} / \alpha_1,..\alpha_p) = \left(\frac{1}{\sqrt{2\pi\sigma_0^{2}}}\right)^{N'_0} \exp\left(\frac{-N'_0 \hat{\sigma}_0^{2}}{2\sigma_0^{2}}\right) \quad (3)$$

joint likelihood ratio:

$$\lambda = \frac{p(l/H_0)}{p(l/H_1)} = \frac{\left(\frac{1}{\sqrt{2\pi\sigma_0^{2}}}\right)^{N_0'} \exp\left(\frac{-N_0' \hat{\sigma}_0^{2}}{2\sigma_0^{2}}\right)}{\left(\frac{1}{2\pi\sigma_R^{2}}\right)^{N_R'} \exp\left(\frac{-N_R' \hat{\sigma}_R^{2}}{2\sigma_R^{2}}\right)}$$

H_0 : $\sigma_R^{2} = \sigma_0^{2}, \alpha_R = \alpha_0$ $\quad$ H_1 : $\sigma_R^{2} \neq \sigma_0^{2}$, $\alpha_R \neq \alpha_0$

$$\lambda = \sigma_R^{N_R'} * \sigma_0^{-N_0'} \exp\left(\frac{N_R' \hat{\sigma}_R^{2}}{2\sigma_R^{2}} - \frac{N_0' \hat{\sigma}_0^{2}}{2\sigma_0^{2}}\right) \quad (4)$$

Furthermore, to estimates the variance terms by using the maximum likelihood, we get the log likelihood ratio to be,

$$\ln\lambda = N_R' \ln\hat{\sigma}_R - N_0' \ln\hat{\sigma}_0 + (N_R' - N_0') \quad (5)$$

the series of $\{-\ln\lambda_i\}$(i=1,2,...N-N_0-1) is called statistical variable series.

For threshold h : $\quad -\ln\lambda > h \Rightarrow H_1$

$\qquad\qquad\qquad\quad -\ln\lambda \leq h \Rightarrow H_0$

Maxion [1], calculate h from known data automatically. let $r = -\ln\lambda$

$$\bar{r} = \frac{\sum_{i=1}^{N-N_0-1} r_i}{N - N_0 - 1}, \quad \sigma = \sqrt{\frac{1}{N - N_0 - 1}\left(\sum_{i=1}^{N-N_0-1} (r_i - \bar{r})^2\right)}$$

h is considered as : $\bar{r} + 2\sigma$ or $\bar{r} + 3\sigma$

Once a change is detected, $\{\lambda_i\}$ correspond to i, we can determine which segment of the slide window occurred change. Then, we will determine the failure point in this segment.

The second stage is to diagnose the failure point. With the purpose of determining the failure point in the anomaly segment, we use wavelet transform. By using wavelet transform, we can decompose the observations into approximate coefficients and detailed coefficients. The traffic anomaly may be detected from detailed coefficients. So, the second step of the WGLR algorithm is to calculate the discrete stationary wavelet transform (DSWT) [4]

2.2.2 Discrete stationary wavelet transform (DSWT)

$x(t)$ is an integrabel function ($x(t) \in L^2(R)$, L^2 represents the Hilbert space), R is a real number. $\psi(t)$ is an wavelet function.

$$WT_x(a,b) = \frac{1}{\sqrt{a}} \int x(t) \psi^*(\frac{t-b}{a}) dt = \langle x(t), \psi_{ab}(t) \rangle$$

The formula above is wavelet transform of $x(t)$.

Here $a > 0$, is scale factor, b is offset, $b \in R$.

$$\psi_{ab}(t) \underline{\underline{\Delta}} \frac{1}{\sqrt{a}} \psi(\frac{t-b}{a})$$

scale discrete: $a = a_0^{\ j}$ .Offset discrete:

When $a = a_0^{\ 0}, b = b_0$; when $a = a_0^{\ j}, b = k \cdot a_0^{\ j} \cdot b_0$, $j \in Z$ .Here $a_0 \neq 1$ and b_0 is constant.

$$WT_x(a_0^{\ j}, kb_0) = \int x(t) \psi^*_{a_0^{\ j}, kb_0}(t) dt$$

$$\psi^{*}{}_{a_0{}^j,kb_0}(t)\underline{\underline{\Delta}}a_0{}^{-\frac{j}{2}}\psi(a_0{}^j t-k\cdot b_0),$$
$$k\in Z\,,\quad j\in Z$$

Select $a_0=2$, $b_0=1$, $a=2^0,2^1,\cdots;b=k\times 2^j$, which is called binary wavelet transform[4].

Classical discrete wavelet transform has not the property of stable transition. Whereas Discrete Stationary Wavelet Transform (DSWT) has the advantage of transition is invariable. It is the property that plays an important part in the failure diagnosis.

For original data, DSWT algorithm calculates the wavelet transform for even element and odd element (The length of detailed and approximate coefficients is N/2). Then join the detailed and approximate coefficients cross over, so the length of the detailed and approximate coefficients of the DSWT is N.

For DSWT, selecting $a=2^1$, $a=2^2$.

We calculate the DSWT for series $\{-\ln\lambda_i\}$ to get the detail coefficient series $\{S_i\}$. To set $\{S_i\}$ as statistical variable, and calculate the threshold h , diagnose the failure point according the threshold.

$$\overline{S}=\frac{\sum_{i=1}^{N-N_0-1}S_i}{N-N_0-1}\quad \sigma=\sqrt{\frac{1}{N-N_0-1}\left(\sum_{i=1}^{N-N_0-1}(S_i-\overline{S})^2\right)}$$

h is $\overline{S}+2\sigma$ or $\overline{S}+3\sigma$ For h : $-\ln\lambda>h\Rightarrow H_1$

$$-\ln\lambda\le h\Rightarrow H_0$$

By making use of the property of transition invariable, we can detect the fault point.

2.3 Error Performance Detection Algorithm (EPD)

Similar to WGLR algorithm, the stationary residual process can be simulated as AR model. AR(P) :

$$X_t\ =\ \varphi_1\ X_{t-1}\ +\ \varphi_2\ X_{t-2}\ +\ ...+\ \varphi_p\ X_{t-p}\ +\ \alpha_t$$
(6)

Estimated value of X_t is:

$$\hat{X}_t=\hat{\varphi}_1 X_{t-1}+\hat{\varphi}_2 X_{t-2}+...+\hat{\varphi}_p X_{t-p}$$

Prediction error is: $e(t)=X_t-\hat{X}_t$, $t=1,2,...N$,

Prediction error at time N+1(current moment) is:

$$e(N+1)=X_{N+1}-\hat{X}_{N+1}\qquad(7)$$

Here $e(t)$ is drawn from an N(0, $\sigma_R{}^2$) distribution.

$$\sigma_e{}^2=\frac{1}{N-1}\sum_{i=1}^{N}(e_t-\overline{e}_t)^2=\frac{1}{N-1}\sum_{i=1}^{N}e_t{}^2$$

We can determine a statistical variable ε to form series $\{\varepsilon_i\}$.

$$\varepsilon(t)=X(t)-\hat{X}(t)\,,\quad t=1,2,...N$$

To identify deviant observations, we first transform each observation as follows [6]:

$$\varepsilon_t=\frac{X_t-\overline{X}_t}{\overline{\sigma}_t}\qquad(8)$$

Replace X_t in Eq.(8) with $e(t)$,

Then $\qquad\varepsilon_t=(e_t-\overline{e}_t)/\overline{\sigma}_e,\qquad \overline{e}_t=0$

Thus $\qquad\varepsilon_t=e_t/\overline{\sigma}_e\qquad$, $\qquad$ t=1,2,….N,

$$\varepsilon(N+1)=e(N+1)/\overline{\sigma}_e\qquad(9)$$

Above mentioned formula means that the statistical variable ε at time N+1 is determined by the ratio of prediction error e(t) and $\overline{\sigma}_e$. The major factor that affecting ε is the prediction error e(t) of simulative model. The range of statistical variable ε is limited by the average value of ε and Standard deviation δ (threshold h is $\overline{\varepsilon}\pm 3\delta$):

$$\overline{\varepsilon}=\frac{\sum_{i=1}^{m}\varepsilon_i}{m},\qquad \delta=\sqrt{\frac{1}{m-1}\left(\sum_{i=1}^{m-1}(\varepsilon_i-\overline{\varepsilon})^2\right)}$$

3 Simulation

3.1 Generation of the simulated traffic

In order to verify the accuracy of the algorithm, we test the algorithm by simulating. Firstly, we need to generate network traffic. Due to burst and complicated characteristic, it is difficult to simulate the network traffic data. Usually, we generate the self-similar process to simulate the real traffic data. What the relation between the short memory model and self-similar process. Article[7]has introduced a state space representation for self-similar signals and systems based on scale stationary ARMA models. It means that the self-similar process can be approximated by scale stationary autoregressive models.

- **Self-Similar traffic data generation**

We select the Inverse Fourier Transform algorithm to generate the Fractal Gaussian Noise (FGN) for simple and rapid characteristic. Giving Hurst parameter and the length of the data, we may synthetic the self-similar traffic sample which has FGN power spectrum [8].Anomaly pulse signal is the stochastic data produced by computer.

Fig.1 (a) is the self-similar data of simulating.

3.2 Simulation results of WGLR algorithm

We can acquire the detailed coefficient D by Discrete Stationary Wavelet Transform (DSWT), then

$$\overline{D}=\frac{\sum_{i=1}^{N-1}D_i}{N-1}\qquad \sigma=\sqrt{\frac{1}{N-1}\left(\sum_{i=1}^{N-1}(D_i-\overline{D})^2\right)}$$

threshold h is $\overline{D}\pm 2\sigma$ or $\overline{D}\pm 3\sigma$. We select SN=4.9523db, the number of the inserting anomaly signal is 5. The detecting result shows in Fig.1(c), Fig.1 (d).

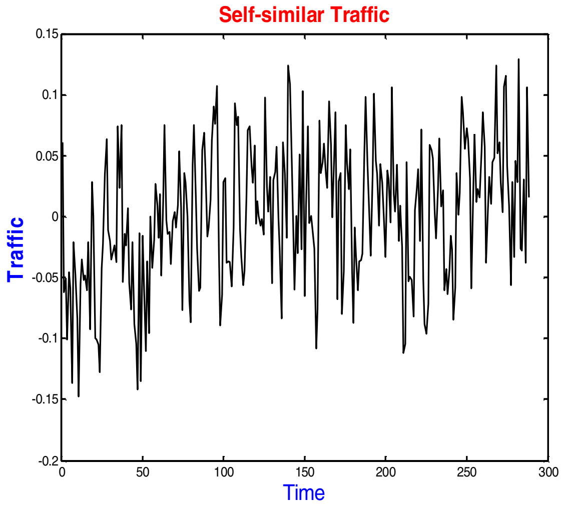


Fig.1 (a) N=288, H=0.8, original data

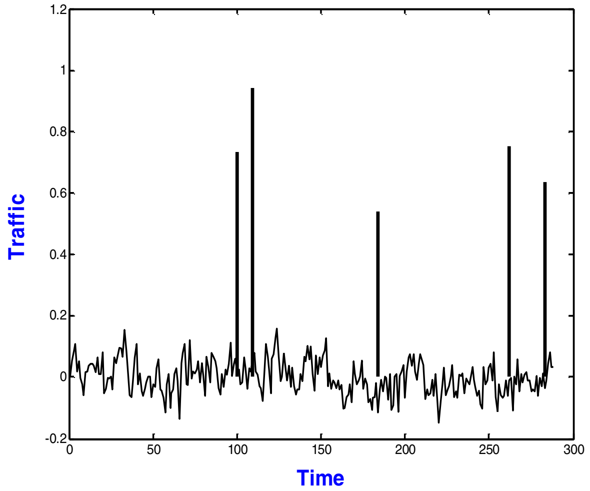


Fig.1 (b) N=288, H=0.8, insert anomaly pulse. SN=4.9523db

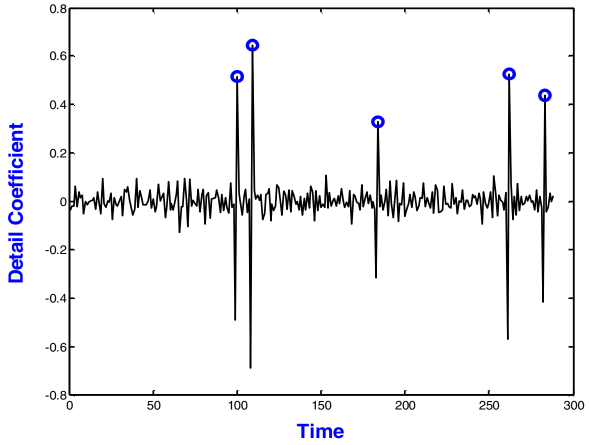


Fig.1(c) Detailed coefficients of Haar wavelet, "o" indicates false alarm

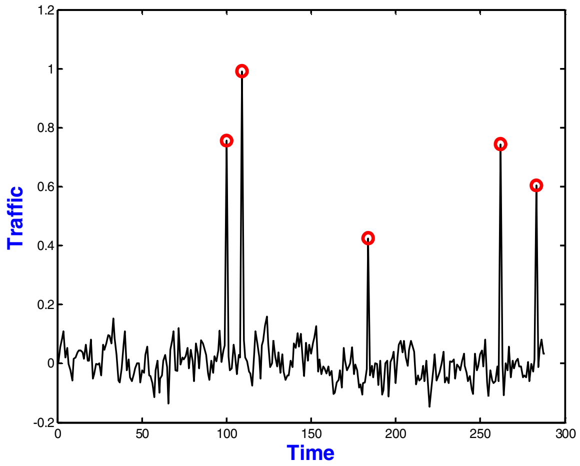


Fig.1(d) N=288, H=0.8, original data detect figure. “o” indicates failure points.

3.3 Results comparison of different algorithm

If we control the amplitude of the signal and noise and make the SN in a fixed DB values, changing the thresholds, we will get the ROC (Receiver operating characteristics) curves at different thresholds. Result shows in Fig.2. N=288, is the number of data samples, the number of known faults is 5(Insert pulse signal)

P_D=number of correct matches/known fault number

P_F=false alarm number/data samples

Table I shows the performance of the different algorithm.

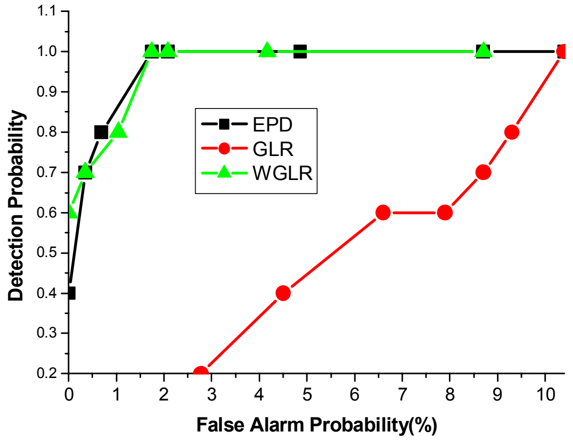


Fig.2 SN=3.2db, results of WGLR, GLR, EPD algorithms.

Table I Performance comparison of different algorithm

method	GLR	EPD	WGLR
Known fault number	5	3	5
Correct match fault number	3	3	5
False alarm number	22	0	2
Detection Probability P_D	0.6	1	1
False alarm P_F （%）	7.64%	0	0.69%

3.4 Analysis of results

The performance of different algorithms can be obtained from ROC curves and Table I. The detection probability P_D of WGLR is higher than GLR and EPD at the same P_F. The P_D of WGLR and EPD could reach 1 at the same SN db, yet the P_D of GLR algorithm could not achieve 1. The P_F of GLR is much higher than WGLR and EPD if all the algorithms have the same P_F.

4 Network Traffic Experiment

4.1 Data Collected

In this work, on the CERNET IP backbone network, we collect values for two particular MIB (Management information Base) objects, incoming and outgoing link utilization in bps, for the links between CERNET and Beijing Internet Centre in the CERNET IP backbone, time spans from January 1st 2004 until June of 2005, time interval is 1 hour. We collect the netflow traffic data at the CERNET international gateway, time interval is 5 minutes. The date of this observation spans from June 6th 2005 to June 19th 2005.

Following figure shows the process for data collecting.

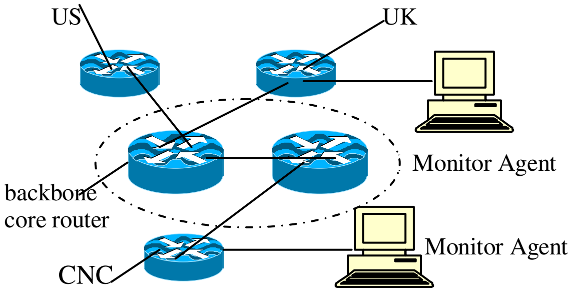


Fig.3 Data collecting system

Fig.4 shows the overall traffic and the 8080 port monitoring. Known faults have been labelled on the figure.

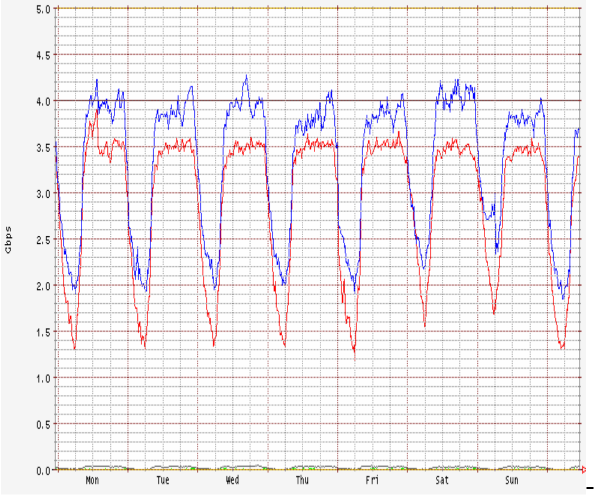


Fig.4(a) week traffic

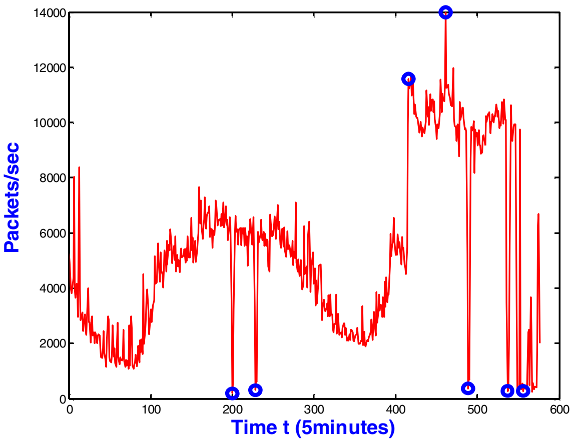


Fig.4 (b) port 8080 traffic
"o" indicates labelled failure points.

4.2 Data Analysis

From the figure has been shown above, we can see that network traffic behaviour exhibits non-stationary property. Regular or periodic seasonal and burst phenomena also appeared. For non-stationary time series, we can also use ANVOA[5] variance analysis method to remove the periodic seasonal component and acquire a stationary time series. Then it is very simple to acquire the autocorrelation of the residual data. The property of the autocorrelation about the residual data is just fit for AR process, so the observed process can be modelled as AR process [6]. Furthermore, MA or ARMA process can be regarded as the AR process when the orders approximate to infinite.

4.3 Experimental Results

To validate the algorithm by network traffic data, we set Autoregressive (AR) model for segment data and execute the WGLR algorithm. Fig.5 is the result of experiment.

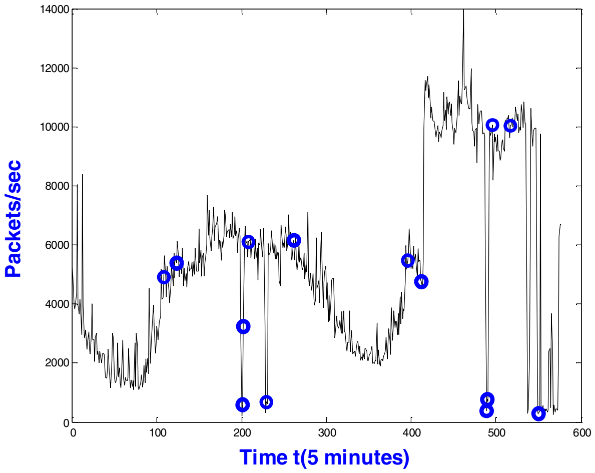


Fig.5 port traffic detecting figure,
"○" indicates detected alarms.

4.4 Performance comparison

The graphic above mentioned obviously shows that the performance of WGLR algorithm. In order to compare the performance of different methods, we present Table II to address the detecting result belongs to different algorithm.

Table II Performance comparison of different algorithm

method	GLR	WGLR	EPD
Known fault number	7	7	7
Correct match fault number	3	6	6
False alarm number	15	4	2
Detection probability P_D	0.4	0.86	0.86
False alarm probability P_F	2.6 %	0.69 %	0.35%

On the basis of experiments, we can see that WGLR algorithm has the advantage of high reliability , and the GLR algorithm has more false alarm numbers. The probability of detection of WGLR is higher and false alarm probability is lower. The performance of WGLR algorithm is superior to GLR . On the basis of experimental results, we can obtain some conclusion.

5 Results And Discussion

5.1 Results of experiment

5.1.1 Reliability of WGLR algorithm is superior to GLR method

A set of experiments, for example, diagnose graph, and detecting precision, which denotes that new algorithm outperforms GLR method. The precision of GLR algorithm is limited by adopting sliding window, the length of the window, which will cause the delay of determining fault point, thus generate the error. Whereas, our new algorithm considers the current time of failure and execute the algorithm in time. Certainly, the computation cost is the least and the calculation is very simple.

5.1.2 Flexibility of the threshold selecting for new algorithm

Our new algorithm generates threshold automatically according to computational results, not manual setting, which will make the algorithm posses the property of flexibility to adapt various data and network environment. Furthermore, the experiment results show that this method does not reduce the reliability for not selecting threshold manually.

6 Summary

In this paper, Wavelet GLR algorithm and EPD method are suggested to solve the network traffic anomaly detection problem. The principle of algorithm has been discussed. Also, the simulation and network experimental result has been presented. Consequently, this paper shows some contributions:

1. Wavelet GLR algorithm is simplified in computation. Only one slide window is to be considered, which minus the number of the test windows. This will reduce two-thirds of the computation cost. Accurately, the CC (computation cost) of AR process is 8N, then the CC of GLR method is 24N, while the CC of WGLR method is 10N, the CC of EPD is 8N.
2. The accuracy of the algorithm has been demonstrated by simulation and network traffic experiment. The results show that the new algorithm has the property of high reliability. WGLR algorithm smartly combines Generalized Likelihood Ratio algorithm and wavelet method, which will detect the failure occurred first and then capture the failure point in real time. All these will improve the performance in diagnosis precision and suggest the feasibility of our algorithm to larger heterogeneous networks.
3. Wavelet transform has a strong ability of detecting abrupt failure points. It could extraction the transient property of the signal in short time range, whereas FFT does not have the ability of capturing the characteristic at local space and does not determine the distribution about the failure points.
4. By applying the algorithm to CERNET port traffic monitoring and detecting, we could detect individual port failure which can not be diagnosed from overall traffic monitoring.

Reference

[1] Maxion Roy A : Anomaly detection for diagnosis, in Proceedings of the 20th International Symposium Fault-Tolerant computing(FTCS-20),1990.20-27

[2] Jun Jiang, Symeon Papavassiliou: A network Fault Diagnostic Approach Based on a Statistical Traffic Normality Prediction Algorithm[C]. Proceedings of IEEE Globecom 2003, 2918-2922.

[3] Thottan Marina, Chuanyi Ji, :Adaptive Thresholding for Proactive Network Problem Detection, IEEE International workshop on systems Management, Newport, Rhode Island, 1998.108-116

[4] Mallat S.A wavelet tour of signal processing [M]. Second Edition. China Machine Press, 2002.

[5] Yang Wei-Qing, Gu Lan: Time series analysis and dynamic data modelling[M]. Press of Beijing University of Science and Engineering, 1988

[6] M.D.Srinath and P.k.Rajasekaran, :An Introduction to Statistical Signal processing with Application,1979

[7] I.Meltem,Y.Birsen,O.Banu: Kalman filtering for self-similar processes. Proceedings of the 11th IEEE Workshop on Statistical Signal processing, Aug. 2001, P.82-85

[8] Paxson V.Fast approximate synthesis of fractional gaussian noise for generating self-similar network traffic [J]. Computer communication review.1997(10):5—18.