Research on Network Traffic Anomaly Detection Algorithm

Jun Lv[1] Tong Li[1] Xing Li[2]
[1]The Academy of Armoured Forced Engineering, Beijing, China
Beijing 100072, China
[2]China Education and Research Network, Tsinghua University,
Beijing 100084, China

Abstract– **Network traffic anomaly detection is a difficult problem in network management. This paper proposed the Wavelet Generalized Likelihood Ratio (WGLR) algorithm and the Error Performance Detection (EPD) algorithm to solve this problem. The WGLR algorithm combines the Generalized Likelihood Ratio (GLR) algorithm and the Wavelet transform method, and has an advantage of less computation cost. The EPD algorithm, which is based on the prediction error of the traffic model, will detect the anomalous change in the signal without test window delay. Simulating and network traffic experimental results show that the new algorithm has the better performance in network traffic anomaly detection.**

1 Introduction

Internet is a large-scale complicated system. Due to the dynamic property of the network, 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.

- The threshold detection approach is preset a threshold by historical data, and compare the current data to this threshold[1]. If the current data exceeds the threshold, the alarm will be generated.
- GLR (Generalized Likelihood Ratio) is commonly used in network problem detection [2]. 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 the data value exceeds the threshold is regarded as anomaly point.
- Jun Jiang[3] proposed a prediction method to detect the network server performance.

In this paper, we propose two new algorithms, Wavelet Generalized Likelihood Ratio (WGLR) algorithm and the Error Performance Detection (EPD) algorithm ,to detect the network anomaly. By monitoring the port traffic and execute the new algorithms, we could be able to find and even predict some port traffic anomaly which can not be detected by threshold method, as well as can not be detected from the overall network traffic.

To verify the effect of the algorithms, we simulate the new algorithm by the computer and experiment it on the China Education and Research network (CERNET). Simulation and the experimental results show that the new algorithm has an advantage of high accuracy with the less computation cost. It determines the anomaly point to a nicety. The simplicity and reliability of the algorithm make it suitable for online implementation.

This paper has been arranged as follows: Section 2 discusses anomaly detection algorithm. InSection 3, Section 4 we will simulate and implementation the new algorithms, 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

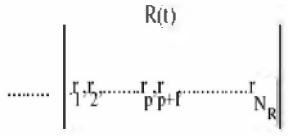
Considering GLR algorithm has some shortages as follows:
(1) Too much test window, complicated computation.
(2) Anomaly detection is not real time, has a delay of the test window, which will affect the accuracy of the detection.
To resolve these problems, we propose the wavelet GLR algorithm (WGLR) and the Error Performance Detection (EPD) algorithm.

2.1 Wavelet GLR Algorithm (WGLR)

The WGLR algorithm is divided into two steps: the first step is calculating the likelihood ratio and detecting the abnormal point.; the second step is to diagnose the abnormal point.

2.1.1 Calculate the likelihood ratio

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



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

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

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

$$\varepsilon_i(t) = \sum_{k=0}^{p} \alpha_k \tilde{r}_i(t \quad k) \qquad (1)$$

where $\alpha_k = \{\alpha_1, \alpha_2, \ldots \alpha_p\}$ 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

$$p(\varepsilon_{p+1}, .., \varepsilon_{N_R} / \alpha_1, .. \alpha_p) = \frac{1}{\sqrt{2\pi\sigma_R^2}}^{N'_R} \exp \frac{N'_R \sigma_R^2}{2\sigma_R^2} \qquad (2)$$

From Equation (2) and assuming that each sample of the residual error $\varepsilon_i(t)$ is drawn from an $N(0, \sigma_R^2)$ distribution[7],

1-4244-1521-7/07/$25.00 ©2007 IEEE

where σ_R^2 is the variance of the residual in segment R(t), and $N'_R = N_R \quad p$ and σ_R^2 is the covariance estimate of σ_R^2. 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. The normal mean and variance were computed from the normal data (over a twenty four hours)

For H_0 :

$$p(\varepsilon_{p+1},..,\varepsilon_{N_0}/\alpha_1,..\alpha_p) = \frac{1}{\sqrt{2\pi\sigma_0^2}}^{N'_0} \exp\frac{N'_0\sigma_0^2}{2\sigma_0^2}$$

(3)

joint likelihood ratio:

$$\lambda = \frac{p(I/H_0)}{p(I/H_1)} = \frac{\frac{1}{\sqrt{2\pi\sigma_0^2}}^{N_0'} \exp\frac{N_0'\sigma_0^2}{2\sigma_0^2}}{\frac{1}{2\pi\sigma_R^2}^{N_R'} \exp\frac{N_R'\sigma_R^2}{2\sigma_R^2}}$$

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

$$\lambda = \sigma_R^{N_R'} \quad \sigma_0^{N_0'} \exp\frac{N_R'\sigma_R^2}{2\sigma_R^2} \quad \frac{N_0'\sigma_0^2}{2\sigma_0^2} \quad (4)$$

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

$$\ln\lambda = N_R' \ln\sigma_R \quad N_0' \ln\sigma_0 + (N_R' \quad N_0') \quad (5)$$

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

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

$\ln\lambda \quad h \quad H_0$

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

$$\bar{r} = \frac{\sum_{i=1}^{N \quad N_0 \quad 1} r_i}{N \quad N_0 \quad 1},$$

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

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

Once a change is detected, $\{\lambda_i\}$ correspond to i, which segment of the slide window occurred change will be determined. Then the abnormal in this segment can be determined.

2.2.2 Discrete stationary wavelet transform (DSWT)

The second stage is using the wavelet transform to diagnose the abnormal point in the anomaly segment. Wavelet transform has a strong ability of detecting abrupt anomaly points. It could extraction the transient property of the signal in a short range of time, whereas FFT does not have the ability of capturing the characteristic at local space.

By using the discrete stationary wavelet transform (DSWT) [4], we can decompose the observations into approximate coefficients and detail coefficients. The traffic anomaly may be detected from detail coefficients. So, the second step of the WGLR algorithm is to calculate the discrete stationary wavelet transform (DSWT).

Support $x(t)$ is an integrabel function , $\psi(t)$ is an wavelet function.

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

$$\psi_{ab}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t \quad b}{a})$$

scale discrete: $a = a_0^j$ .offset discrete:

when $a = a_0^0$, $b = b_0$; when $a = a_0^j$, $b = k \quad a_0^j \quad b_0$, $j \quad Z$. here $a_0 \quad 1$ and b_0 is constant.

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

$$\psi^*_{a_0^j, kb_0}(t) = a_0^{\frac{j}{2}}\psi(a_0^j t \quad k \quad b_0), k \quad Z \ , \ j \quad Z$$

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

Discrete Stationary Wavelet Transform (DSWT) calculates the wavelet transform for even element and odd element (The length of detail and approximate coefficients is N/2). Then join the detail and approximate coefficients cross over, so the length of the detail and approximate coefficients of the DSWT is N. Therefore, DSWT has the characteristic of transition invariable. It is this property that plays an important part in the failure diagnosis.

For DSWT, usually selecting one or two layers to do wavelet discompose, therefore select $a = 2^1$ (one layer) , $a = 2^2$ (two layer).

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 anomaly point according the threshold.

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

h is $\bar{S} + 2\sigma$ or $\bar{S} + 3\sigma$.

For h : $\ln\lambda > h \quad H_1$, $\ln\lambda \quad h \quad H_0$

By making use of transition invariable characteristic, we can

detect the abnormal point.

2.3 Error Performance Detection Algorithm (EPD)

In order to reduce the computation cost, We also propose EPD algorithm. The stationary residual process can be simulated as AR model. AR(P) is :

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

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

Prediction error: $e(t) = X_t - X_t$, $t = 1,2,...N$,

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

$e(t)$ 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 compose series $\{\varepsilon_i\}$.

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

To identify deviant observations, we first transform each observation as follows [7]:

$$\varepsilon_t = \frac{X_t - \overline{X_t}}{\overline{\sigma_t}} \qquad (8)$$

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

Then $\varepsilon_t = (e_t - \overline{e_t})/\overline{\sigma_e}$, $\overline{e_t} = 0$

Thus $\varepsilon_t = e_t/\sigma_e$ t=1,2,....N,

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

$$\overline{\varepsilon^+} = \frac{\sum_{i=1}^{m}\varepsilon_i}{m}, \quad \overline{\varepsilon^-} = \frac{\sum_{i=1}^{n}\varepsilon_i}{n}$$

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

$$\delta^- = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n-1}(\varepsilon_i - \overline{\varepsilon^-})^2}$$

The limit range is $(\overline{\varepsilon^-} - 3\delta^-, \overline{\varepsilon^+} + 3\delta^+)$.

2.4 computation cost

WGLR 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 (CC). Accurately, the CC 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.

3 Simulation

3.1 Generation of the simulated traffic

Usually, we generate the self-similar process to simulate the real traffic data. The self-similar process can be express by short memory model [5]. Article [5] has introduced that 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**

By using the Inverse Fourier Transform algorithm the Fractal Gaussian Noise (FGN) for simple and rapid property has been generated. Giving Hurst parameter and the length of the data, we may synthetic the self-similar traffic sample which has FGN power spectrum [6].Anomaly pulse signal is the stochastic data produced by computer. Fig.1 (a) is the self-similar data of simulating.

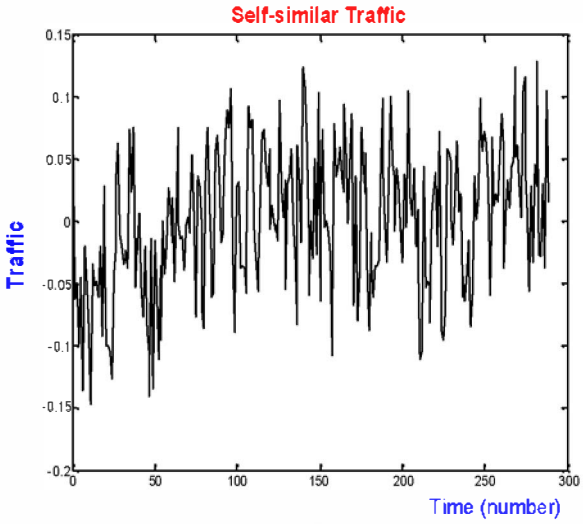


Fig.1 (a) N=288, H=0.8, original data

3.2 Simulation of WGLR algorithm

If D is the detail coefficients decomposed by wavelet transform, then

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

threshold h is $\overline{D} - 2\sigma$ or $\overline{D} - 3\sigma$.

The ratio of the signal and noise is S/N, we select S/N=4.9523, the number of the inserting anomaly signal is 5 ,as show in fig1 (b). The detecting result shows in Fig.1 (c), Fig.1 (d).

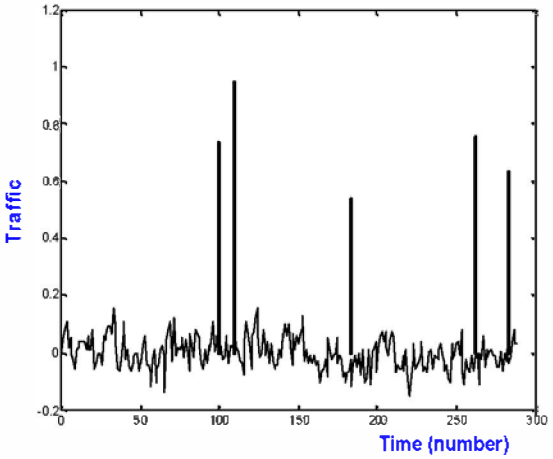


Fig.1 (b) insert anomaly pulse, N=288, h=0.8,S/N=4.9523db

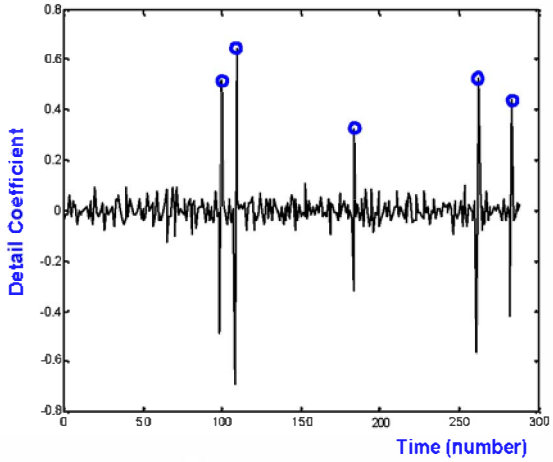


Fig.1(c) Detail coefficients of Haar wavelet, "o" indicates false alarm

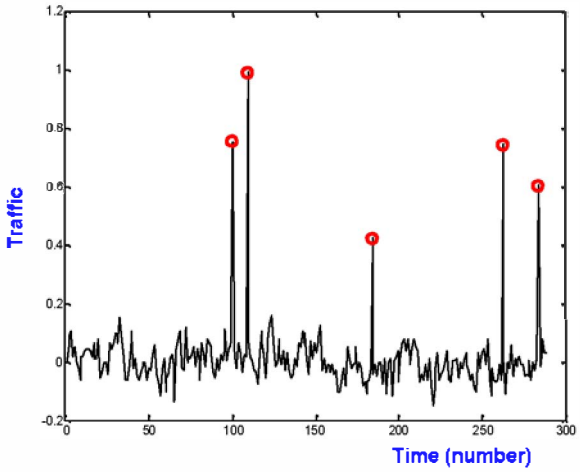


Fig.1(d) N=288, H=0.8, original data detect figure. "o "indicates anomaly points.

3.3 Results comparison of these algorithm

A. Check the relation between P_D and P_F

Let the S/N in a fixed values, changing the thresholds, we will get the ROC (Receiver operating characteristics) curves at different thresholds. Result shows in Fig.2., the total number of data samples(N) is 288; the total number of known faults is 5 (Insert pulse signal) Table1 shows the result of the different algorithm

P_D = number of correct matches/known fault number

P_F = false alarm number/data samples

Table1 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%

Fig.2(a), Fig.2(b),Fig.2(c) shows the ROC curves of EPD, WGLR, GLR algorithm. Fig.2(d) is the results comparing of these algorithm. (N=288, H=0.8, SN=3.2db)

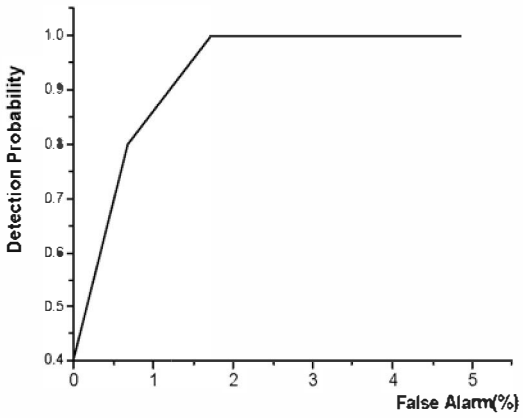


Fig.2 (a) ROC Curve of EPD algorithm, SN=3.2db

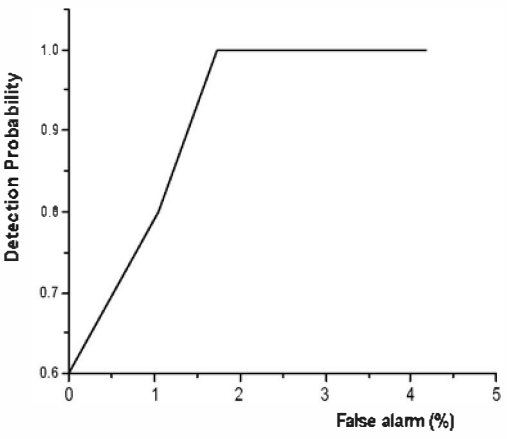


Fig.2 (b) ROC Curve of WGLR algorithm, SN=3.2db

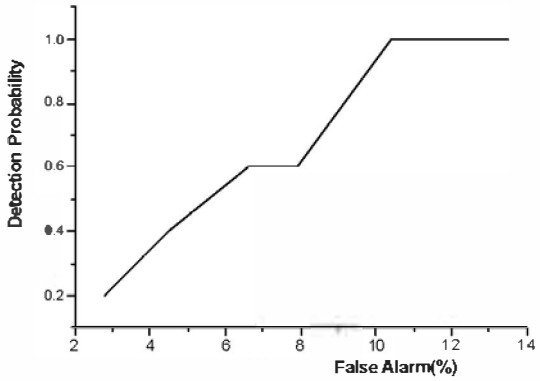


Fig.2 (c) ROC Curve of GLR algorithm, SN=3.2db

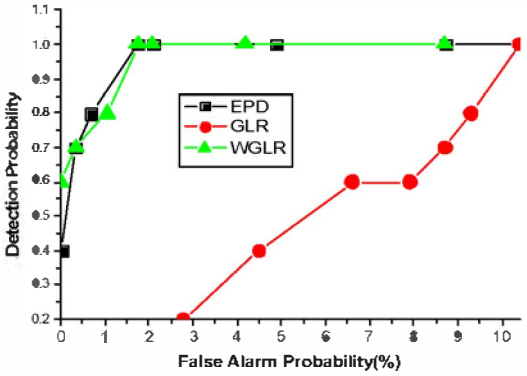


Fig.2(d) S/N=3.2db, results of WGLR, GLR, EPD algorithms.

B. Check the relation between P_D and S/N

Let the false alarm probability P_F in some range (P_F<0.5), changing the rate of signal and noise (S/N),we will get the detecting result at different S/N value. showing in Tabel 2.

Table 2 Detecting result of WGLR algorithm (N=288,H=0.8, P_F <0.5)

Times	1	2	3	4	5
SN(db)	1.4	2.6	3.1	3.7	4.4
Known fault number	4	5	4	5	4
Correct match fault number	1	2	3	4	4
Detecting Probability(P_D)	0.25	0.4	0.75	0.8	1

3.4 Analysis of results

The performance of different algorithms can be obtained from ROC curves Fig.2(a) ~ Fig.2(c), Table 1 and Table2. 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 S/N. The P_F of GLR is much higher than the P_F of WGLR and EPD under the same P_D.

4 Network Traffic Experiment

4.1 Data Collected

We use CERNET as the test network .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 traffic of the CISCO router., which has been collected at the international gateway. Time spans April 2005 to June, time interval is 5 minutes.

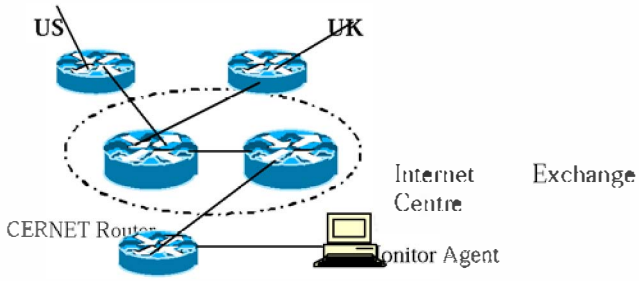


Fig.3 Data collecting system

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

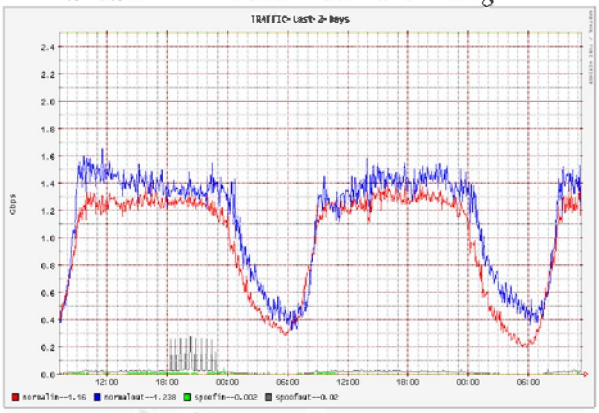


Fig.4(a) two days traffic

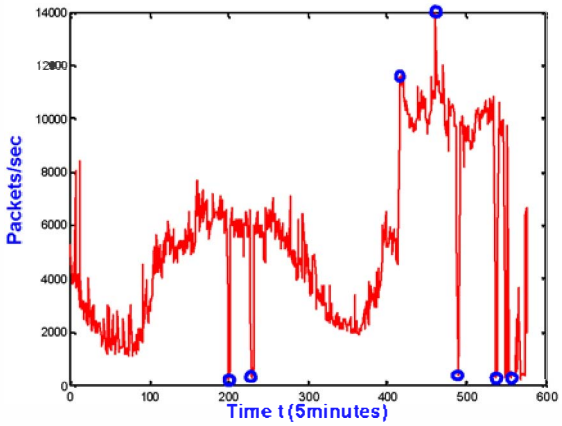


Fig.4 (b) port 8080 traffic
"o " indicates labelled anomaly points.

4.2 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 results of experiment.

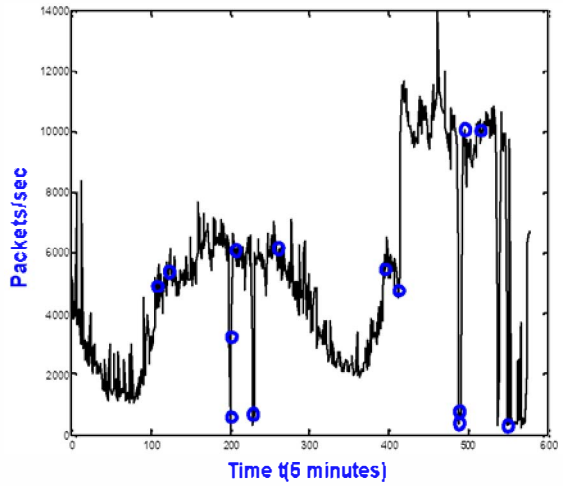


Fig.5 port traffic detecting figure, "○" indicates detected alarms.

4.3 Performance comparison

The performance of these algorithms conclude as Table 3

Table 3. Performance comparison of different algorithm

Algorithm	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 .

5 Discussion and conclusion

A WGLR Algorithmic detecting function is better than GLR

A set of simulation and experiments above, for example, ROC curve, diagnose graph, and detecting precision, which denotes that new algorithm is superior to GLR method. The precision of GLR algorithm is limited by adopting sliding window, the length of the window, which will cause the judgment delay of the anomaly point. Whereas, our new algorithm can find the current time of the network failure, and execute the algorithm in real time.

B The effect of self-similar traffic Hurst parameter

The probability of detection has a little relation with Hurst parameter. Detecting probability P_D changes less when Hurst parameter increases or falls.

6 Summary

In this paper, The WGLR and EPD algorithm is proposed to solve the network traffic anomaly detection problem. Comparing with the GLR algorithm, WGLR and EPD algorithm has the characteristic of high reliability with the less computation cost. The simulation and network experimental results have show that the new algorithm has the property of high reliability. Moreover, by replacing manual setting, the threshold can be automatically calculated from historical data. All these will improve the performance in diagnosis and suggest the feasibility of our algorithm to larger heterogeneous networks.

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

[3] 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.

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

[5] 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

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

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