2009 Fifth International Conference on Information Assurance and Security

Network Traffic Anomaly Detection Based on Self-Similarity Using HHT and Wavelet Transform

Xiaorong Cheng, Kun Xie, Dong Wang
School of Computer Science and Technology
North China Electric Power University
Baoding, China
eric.pat@163.com

***Abstract*—Network traffic anomaly detection can be done through the self-similar analysis of network traffic. In this case, the abnormal condition of network can be indicated by investigating if the performance parameters of real time data locate at the acceptable ranges. A common method of estimating self-similar parameter is the Wavelet transform. However, the Wavelet transform fails to exclude the influence of non-stationary signal's periodicity and trend term. In view of the fact that Hilbert-Huang Transform (HHT) has unique advantage on non-stationary signal treatment, in this paper, a refined self-similar parameter estimation algorithm is designed through the combination of wavelet analysis and Hilbert-Huang Transform and a set of experiments are run to verify the improvement in the accuracy of parameter estimation and network traffic anomaly detection.**

Keywords-anomaly detection; HHT; EMD; wavelet transform; self-Similar

I. INTRODUCTION

Most of the traditional anomaly detection methods are based on feature matching, and its accuracy needs to be improved. Network traffic's self-similar parameter estimated through sampling detection and fast algorithm can judge whether the present network data stream is in accordance with self-similarity and its features, if not, so the present net is abnormal.

Hurst parameter (H) is paramount to describe network traffic's long-range dependence. If 0.5<H<1, sequence has long-range dependence, if not, sequence is abnormal. The real-time and accuracy of Hurst parameter estimation algorithm is a must. Although there are many methods on estimation [2], in actual cases, different methods could make big difference on both the results and time complexity. Therefore, it is important to know the performance of each method and the factors that affects the accuracy. The present typical estimation algorithm of Hurst is using wavelet transform. The multi-scale invariance of wavelet has natural connection with self-similar process, it is multi-resolution dominant on fractal signal processing and fractal parameter estimation, and it has higher precision and speed. However, as a real-time and random time series, the network traffic detection signal has periodicity, which will make some effects on wavelet transform [3]. The periodicity is bigger, and the inaccuracy is bigger. Besides that, there is no inevitable connection between self-similar process and stationary process. But usually speaking, normal self-similar process is non-stationary random process. This kind of nonstationarity would be seen as the trend term, which would affect the estimation of Hurst parameter [4].

In this paper, a Hurst parameter estimation algorithm which combines wavelet transform with Hilbert-Huang Transform (HHT) is designed to improve the accuracy of the results. HHT is a self-adaptive time frequency analysis method, and it can apply on both non-stationary and nonlinear signals. The effect of periodicity and nonstationarity trend term can be eliminated through HHT to improve estimation accuracy, and at the same time, the accuracy of anomaly detection is also improved.

II. PARAMETER ESTIMATION ALGORITHM

A. Estimation Based on Wavelet Transforms

In this paper, wavelet coefficients variance method is used, which can estimate Hurst parameter form transform coefficients.

Discrete Wavelet Transform (DWT) could split the signal as general picture and detail, viz.

$$\begin{aligned} X(t) &= \text{approx}_J(t) + \sum_{j=1}^{J} \text{detail}_j \\ &= \sum_k a_x(J,k)\varphi_{J,k}(t) + \sum_{j=1}^{J}\sum_k d_x(j,k)\psi_{j,k}(t) \end{aligned} \quad . \tag{1}$$

Among them, $\varphi_{J,k}(t)$ is a low-pass filter, $\psi_{j,k}(t)$ is a band-pass filter.

Decomposition coefficients could be achieved through the inner product of X(t) and $\varphi_{J,k}(t)$ or X(t) and $\psi_{j,k}(t)$, viz.

$$a_x(J,k) = \langle X, \varphi_{J,k} \rangle = 2^{\frac{J}{2}} \int x(t)\varphi(2^J t - k)dt , \tag{2}$$

$$d_x(j,k) = \langle X, \psi_{j,k} \rangle = 2^{\frac{j}{2}} \int x(t)\psi(2^j t - k)dt . \tag{3}$$

If X(t) is a two order self-similar process, having power spectrum $P_x(\omega)$, and

$$P_x(\omega) \propto \frac{1}{|\omega|^{\gamma}}, \gamma = 2H+1 . \tag{4}$$

Make Discrete Wavelet Transform on X(t), and wavelet coefficients $d_x(j,k)$ can be get through (3), in which

978-0-7695-3744-3/09 $25.00 © 2009 IEEE
DOI 10.1109/IAS.2009.219

$\psi_{j,k}(t)=2^{\frac{j}{2}}\psi(2^j t-k)$ is a dyadic orthogonal wavelet, whose regular degree is R. The expectation data of wavelet coefficients is 0, viz.

$$E[d_x(j,k)]=E[X(t)]2^{\frac{j}{2}}\int\psi(2^j t-k)dt=0. \quad (5)$$

According to the definition of correlation coefficient, the correlation degree of two arbitrary wavelet coefficients $d_x(j,k)$ and $d_x(j',k')$ is:

$$E[d_x(j,k),d_x(j',k')]=\iint E[X(t)\psi_{j,k}(t)X(t')\psi_{j',k'}(t')]dtdt'. \quad (6)$$

Make Fourier transform and use Parseval formula, and get:

$$E[d_x(j,k),d_x(j',k')]=\frac{2^{-(j+j')/2}}{2\pi}\cdot\int_{-\infty}^{+\infty}\frac{\sigma_x^2}{|\omega|^\gamma}\hat{\psi}(2^{-j}\omega)\overline{\hat{\psi}(2^{-j'}\omega)}e^{-i(k2^{-j}-k'2^{-j'})\omega}d\omega. \quad (7)$$

According to (6) and (7), the variance is calculated:

$$\mathrm{var}[d_x(j,k)]=\frac{2^{-j\gamma}}{2\pi}\cdot\int_{-\infty}^{+\infty}\frac{\sigma_x^2}{|\omega|^2}|\hat{\psi}(\omega)|^2\,d\omega. \quad (8)$$

Then can get:

$$\sigma^2=\frac{1}{2\pi}\cdot\int_{+\infty}^{-\infty}\frac{\sigma_x^2}{|\omega|^2}|\hat{\psi}(\omega)|^2\,d\omega. \quad (9)$$

According to (8) and (9), then can get:

$$\mathrm{var}[d_x(j,k)]=\sigma^2 2^{-j\gamma}. \quad (10)$$

Get logarithm from two sides of (10). Make linear fitting and make sure the mean square error is least. Then a straight line can be made with independent variable j and $\log_2\mathrm{Var}[d_x(j,k)]$ as its function, and its slope is γ. The estimated value of H can be calculated from $\gamma=2H+1$.

B. Pretreatment Based on HHT

HHT has two steps. First, split the complicated signal into several Intrinsic Mode Functions (IMF) by Empirical Mode Decomposition (EMD). Then make Hilbert Transform to get Hilbert spectrum, and get the marginal spectrum.

In this paper, the first step, Empirical Mode Decomposition is mostly used. IMF must satisfy two conditions, firstly extreme points (maximal and minimal value) must equal or one data difference at most with the cross-zero point; secondly the mean value of upper envelope that constituted by local maximum value and lower envelope that constituted by local minimum value must be 0.

The decomposition on network traffic signal through EMD can eliminate the influence of the trend term and periodicity of original signal.

Get the local extreme points of network traffic signal X(t), then connect all the local maximum points to form upper envelope $e_{max}(t)$ and all the local minimum points to form lower envelope $e_{min}(t)$, which should envelope all the data points. The mean value of upper and lower envelope is m(t) , viz.

$$m(t)=\frac{e_{max}(t)+e_{min}(t)}{2}. \quad (11)$$

X(t) minus m(t) is the remainder $h_1(t)$,

$$h_1(t)=X(t)-m(t). \quad (12)$$

If $h_1(t)$ can fit the two conditions of IMF, then it can be the first IMF component of X(t). Generally speaking, $h_1(t)$ is not a IMF, due to the nonlinear and nonstationarity of the signal. Therefore, take $h_1(t)$ as the new original data and repeat the steps above, then get the mean value m'(t) of upper and lower envelope to achieve the new remainder to see if the remainder is up to the mustard. If not, continue to cycle until the fit IMF is gotten as $c_1(t)$.

Separate $c_1(t)$ from X(t) and get:

$$r_1(t)=X(t)-c_1(t). \quad (13)$$

Make $r_1(t)$ the new original data and repeat above steps to get the second component $c_2(t)$. Repeat n times to get n IMF components until the stopping criterion of the screening process is meet, viz.

$$\begin{cases} r_2(t)=r_1(t)-c_2(t) \\ \vdots \\ r_n(t)=r_{n-1}(t)-c_n(t) \end{cases}. \quad (14)$$

From (13) and (14), can get:

$$X(t)=\sum_{j=1}^{n}c_j+r_n. \quad (15)$$

So network traffic can be separated as the total of n basic mode components and a residual function r_n. Component $c_1,c_2\ldots c_n$ contains the ingredient from high-frequency to low-frequency of the signal. r_n contains trend information of original signal as the trend term.

The modulus of IMF component is not less than 7 [5] for any self-similar signal, so in this paper chose the max modulus 7, which means to get 7 IMF.

Based on the pretreatment of algorithm above, the self-similar information is still remains in original signals, and the trend term r_n can be eliminated from network traffic signal to eliminate the influence of the trend term and periodicity and to improve the estimation accuracy.

C. Application Process of the Algorithm

First, do the HHT pretreatment. Then estimate H based on wavelet transform. Finally judge the validity of H to detect abnormity. The details of process are like Fig.1:

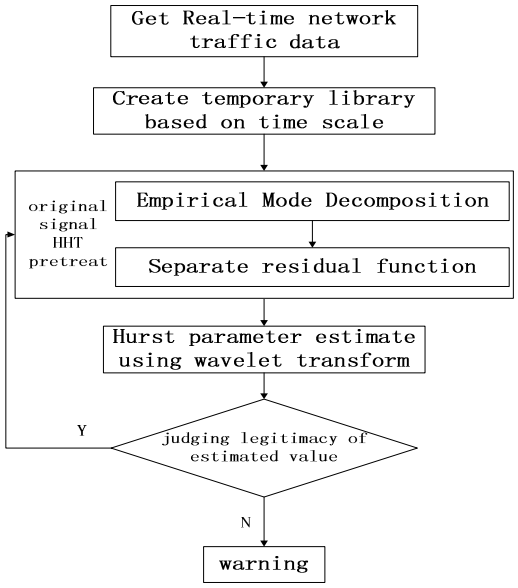

Figure 1. Application process of the algorithm

III. EXPERIMENT AND ANALYSIS

A. *Analysis of the Two Different Algorithms*

The two algorithms mean using wavelet transforms alone and using the algorithm combining HHT and wavelet transform. For comparison, data sequences with different H are chosen and use these two algorithms to estimate their H at the same time. Since the value of H is already known, comparing the size of error can prove which algorithm is good.

Fig. 2 gives 4 data sequence with different H.

The estimation results obtained from these two algorithms are shown in table 1.

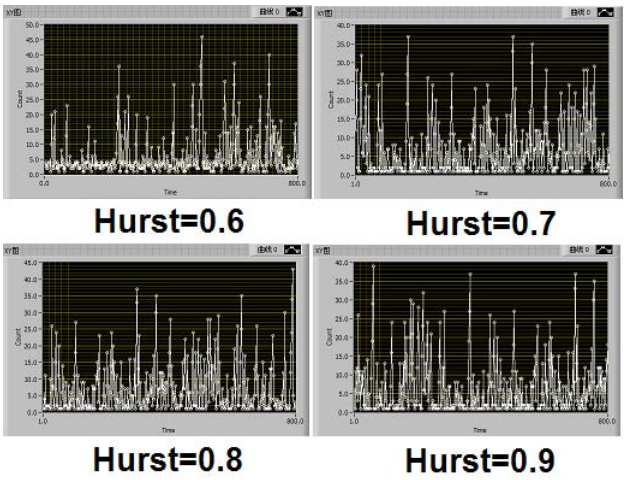

Figure 2. Data sequence in different Hurst parameter

TABLE I. HURST PARAMETER ESTIMATE VALUE

Hurst parameter estimate	Different Hurst			
	H = 0.6	***H = 0.7***	***H = 0.8***	***H = 0.9***
Wavelet analysis	0.6375	0.7382	0.8464	0.9359
Wavelet & HHT	0.6208	0.7191	0.8213	0.8987

As shown in table 1, the error of the algorithm combining HHT and wavelet transform is clearly smaller than the error of wavelet transform, which proves that making a HHT pretreatment can improve the H estimation accuracy.

According to the self-similar theory, network anomaly detection is based on the estimation of H. Using the algorithm combining HHT and wavelet transform, the detection rate will improve obviously compared with wavelet transform.

B. *Experiment of the Detection Algorithm*

1998 DARPA intrusion detection evaluation data is the world's first Intrusion Detection standard data set, of which the network data is called KDDcup'99 data set. In this Paper, corrected.gz data set is used as the experiment data. The reason why choose this data set is to facilitate the validation of the method and prove its effectiveness. The data content used is the No. 23 field, which expresses the amount of data flow connected to the same host in the past 2s.

Fig. 3 shows the data flow in the pre-1600s under normal circumstances.

Select a time scale (200), pretreat the real-time data using HHT, then the use wavelet transform to estimate H. The results are shown in Fig. 4. At the beginning of 1600s, estimated values of H are between 0.5 and 1, which shows the self-similarity of normal network traffic.

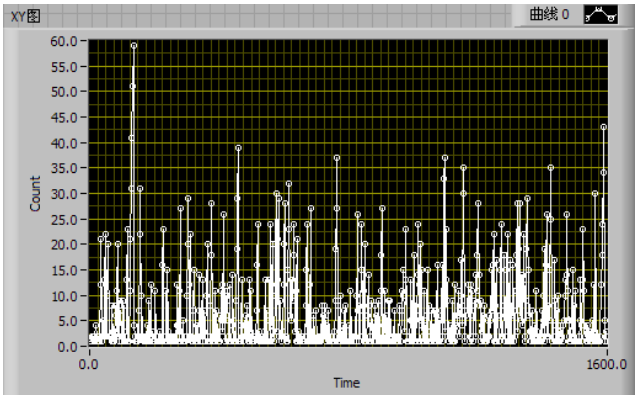

Figure 3. Normal network traffic

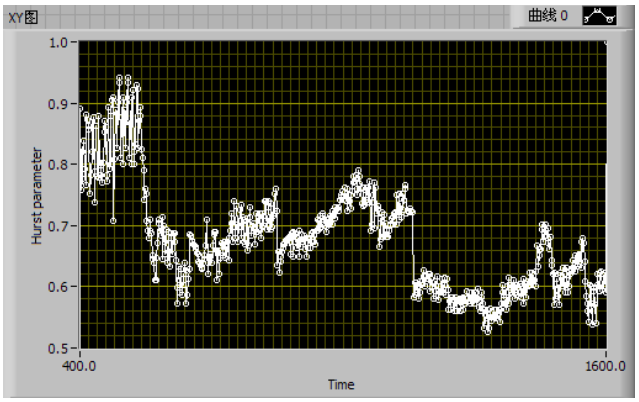

Figure 4. Hurst parameter of normal network traffic

According to the experiment data, Smurf attacks occur at the time around 1600s, which is shown in Fig. 5.

Estimated values of H are shown in Fig. 6.

According to the Fig. 6, the parameter's change is clear. The highest value of H is near to 2.3 which overstep the H's normal value. At about 1600s, H's estimated value is bigger than 1, which is the maximum in normal condition. The network abnormity (flooding attacks) is found in time, which validates the effectiveness of the method.

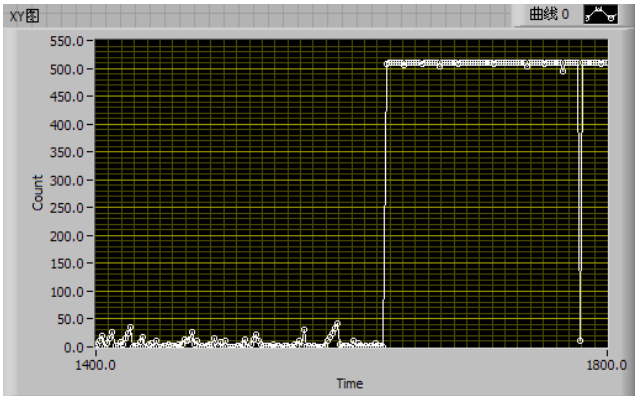

Figure 5. Network traffic under Smurf attack

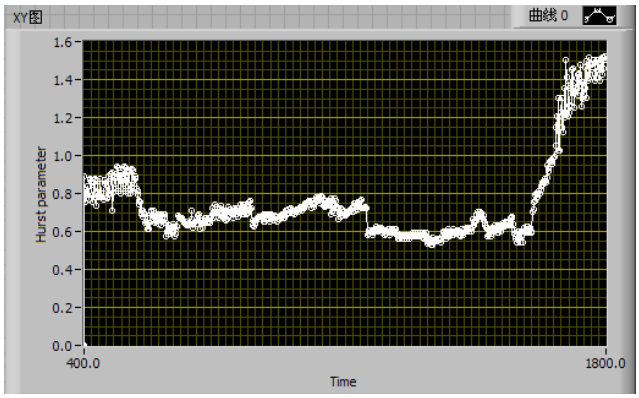

Figure 6. Hurst parameter of network traffic under Smurf attack

IV. CONCLUSION

In this paper, a refined Hurst parameter estimation algorithm is designed and its real-time feasibility and effectiveness are examined. The results of experiments has proved that the algorithm combining HHT and wavelet transform is equipped with better accuracy in estimation and detection rate in network traffic anomaly detection and has made distinct effect in eliminating the impact of periodicity and trend term.

REFERENCES

[1] SHAN Peiwei, LI Ming. "Estimation of Hurst Index of Self-similar Traffic Based on EMD", Computer Engineering, 2008, 34(23). pp. 128-129.

[2] Chen Jian, Tan Xianhai, and Jia Zhen, "Performance analysis of seven estimate algorithms about the Hurst coefficient", Computer Applications, 2006, 26(4). pp. 945-947.

[3] LIN Qingjia, CHEN Di, and LIU Yuncai, "Performance analysis of the estimate algorithms about network traffic's long-range dependence characteristics", Journal of ShanDong University, 2005, 40(1). pp. 86-89.

[4] Dang Trang Dinh, Molnar S, "On the Effects of Non-stationary in Long-range Dependence Test", Periodica Polytechnic Electrical Engineering, 1999, 43(4). pp.227-250.

[5] Li Ming, "Change Trend of Averaged Hurst Parameter of Traffic under DDOS Flood Attacks", Computers & Security, 2006, 25(3). pp. 213-220.