doi:10.2498/cit.1002007

Network Traffic Deviation Detection Based on Fractal Dimension

Afshin Shaabany and Fatemeh Jamshidi

Science and Research Branch, Islamic Azad University, Fars, Iran

In this paper we examine aggregate network traffic for deviation detection. The precise and fast detection of network traffic deviation is crucial to improve the efficient operation of a network. It is often difficult to detect the time when the defects occur in a network. In this article, a new algorithm is presented to supervise the aggregate network traffic to fast detect the time deviation transpires in a network. This is performed by supervising the statistical attributes of the time series representing the network conduct. The method examines the network conduct using fractal dimension and discrete stationary wavelet transform. In the proposed method, after applying discrete stationary wavelet transform on the signal representing the network traffic, the fractal dimension of the disintegrated signal is computed in a sliding window. Then, variations of signal fractal dimension are considered for deviation detection. Performance of the proposed method is compared with that of three other existing methods using synthetic signal. The results show superiority of the proposed method in terms of accuracy compared to existing methods.

Keywords: deviation detection, network traffic conduct, fractal dimension, detection accuracy

1. Introduction

The network traffic deviation pertain to the condition when network traffic leaves from its normal conduct [1-3]. Deviation detection methods attempt to provide normal activity by regarding different quantification, and a deviation is detected once the actual system conduct deviates from the normal profiles. Many reasons such as defective components, their transient failures, software failures, invasion of the network, and misuse of network equipment may cause network traffic deviation conducts [1-5]. Fast and precise detection of network traffic deviation is very important in network management and improvement of network performance [1-3]. It is often difficult to implement these goals; therefore detection of network traffic deviation has become the interesting and important subject in the present academic and industrial research. Signal processing methods can be used to examine and detect network deviations due to their potential to find novel or unknown deviations. Deviation detection algorithms are based on the assumption that the statistical attributes of the Management Information Base parameters change in reply to defect incident [1-3].

Deviation detection in a network is either host-based or network-based [1-5]. The former systems run on a local supervised host and use its log files as information sources. This deviation detection system has limitations in detecting scattered and matched attacks. In contrast, network-based deviation detection tries to protect the entire networks versus violation by supervising the network traffic either on designated hosts or particular sensors; therefore, it can protect concurrently a group of computers employing different operating systems versus remote attacks such as port scans.

In this article, we present a deviation detection algorithm based on discrete stationary wavelet transform (DSWT) and fractal dimension (FD). Wavelet method has been used to exhibit the important underlying unadulterated form of the time series. This preprocessing step is used to increase the accuracy of the proposed method in deviation detection. The main advantage of discrete stationary wavelet transform is the preservation of time information of the original signal sequence at each level, compared to classical wavelet transform such as proposed in [6]. In the next step, the fractal dimension of the disintegrated signal is computed. For application of deviation detection, fractal dimension variations can be used, as the change in statistical attributes of a signal affect its corresponding fractal dimension. Network traffic experiments have demonstrated that the proposed method is efficient in deviation detection. In addition, a comparative study between four typical wavelet basis functions on the proposed method accuracy has been performed when applying wavelet methods for detecting network deviation. The method is network-based. In a network topology nodes are connected to a LAN or other network via links.

There are a number of deviation detection methods in the literature. In generalized likelihood ratio (GLR) method [3], two successive windows, namely *R* and *S*, are sided along the signal.

$$R = \{r_1,\ r_2,\ \cdots,\ r_p,\ r_{p+1},\cdots,\ r_{N_R}\} \quad (1)$$

$$S = \{s_1,\ s_2,\ \cdots,\ s_p,\ s_{p+1},\cdots,\ s_{N_R}\} \quad (2)$$

where N_R represent the length of the sliding windows. For each window, the observations are modeled as autoregressive (AR) order *p* process. The following function, namely *G* function, is used to detect deviation points:

$$G = N_R^* \left(\ln \frac{\hat{\sigma}_p}{\hat{\sigma}_R}\right) + N_S^* \left(\ln \frac{\hat{\sigma}_p}{\hat{\sigma}_S}\right) \quad (3)$$

where $N_R^* = N_R - p$ and $\hat{\sigma}$ denotes variance estimate. Local maxima of *G* function above the threshold value are considered as deviation points.

In [7], the authors proposed a deviation detection method, namely WGLR, which combines wavelet transform and GLR method. The authors have used the Wavelet transform to employ its strong ability in detecting abrupt failure points. It could also extract the transient characteristic of the signal in short time. It has been shown that the accuracy of this method is superior to GLR method. In another research, Error Performance Detection (EPD) method is introduced for deviation detection [7]. This method is based on the prediction error of the traffic model and regards the error as a statistical variable. The points where error values exceed the predefined threshold are regarded as deviation points.

In [1], a deviation detection method has been proposed which detects traffic deviation by computing and analyzing network traffic signal instantaneous parameters (frequency and amplitude) obtained by Generalized Hilbert Transform of original traffic data. It has been revealed that deviation points would be more evident through analyzing the instantaneous parameters of the original network flow data.

2. Fractal Dimension

Fractal dimension value is an index for measuring the complexity of an object [15, 16]. Its applications have been considered in different fields such as criminology, epidemiology, and economy, social and conducted sciences [15, 16]. It has been shown that *FD* is a promising method in transient detection [11-16]. In addition, in this approach there is no need to have a prior knowledge about the attributes of the transient. Several algorithms have been suggested to compute the fractal dimension of waveforms such as Higuchi [13], Petrosian [15] and Katz [12] methods. Selection of *FD* algorithm depends on the application [14]. Higuchi's method presents the most precise estimation of the *FD*, but it is slower than Katz's method. Petrosian's method has less accuracy in calculating *FD* of a signal compared to Katz's method [14]. In this paper, we have used Katz's algorithm to calculate the *FD* of a time series due to its speed and accuracy. In this method, *FD* of a time series is defined as below [16]:

$$FD = \frac{\log(L)}{\log(d)} \quad (4)$$

where *L* is the length of the time series, *d* is diameter estimation of the distance between the first data point and the data with the highest distance. By normalizing the distance, with *a* being the average distance between the two successive data points, the following equation is obtained.

$$FD = \frac{\log\left({}^{L}\!/_{a}\right)}{\log\left({}^{d}\!/_{a}\right)} \quad (5)$$

The above equation is known as Katz's method to calculate *FD* of the time series.

3. Proposed Method

The proposed deviation detection method consists of three steps as described below:

I. The analyzing signal is initially disintegrated into different frequency bands using discrete stationary wavelet transform. The wavelet method has been used to reveal the important underlying utter form of the data since details have been removed during filtering. In this paper, discrete stationary wavelet transform has been used to preserve the time information of the original signal sequence at each level. This is the main advantage of using discrete stationary wavelet transform in pre-processing step instead of using classical wavelet transform, such as proposed in [6].

II. Two successive windows are sided along the signal. For each window, *FD* is computed using the Katz algorithm. As mentioned before, the changes in the statistical attributes of the signal are reflected on the signal fractal dimension. Therefore, for application of deviation detection, variations of signal fractal dimension can be considered as follows [6]:

$$G_k = |FD_{k+1} - FD_k|\,,\ k = 1, \cdots, N \quad (6)$$

where N is the number of samples in the G function.

III. The local maxima of G function that are above the threshold represent the time instants of defect incident. Threshold value can be chosen automatically according to computational results.

In the proposed approach, a criterion has also been employed to choose a proper length for the sliding window which increases the accuracy of the proposed method. For an analyzing window with length l, the energy of the corresponding G function G^l is computed as below:

$$E_{G^l} = \frac{\sum_k |G_k^l|^2}{N} \quad (7)$$

where N is the number of samples in G^l. The proper window length is the minimum point of the normalized energy function, E_{G^l} versus window length [7].

4. Network Traffic Experiment

For the purpose of evaluating the performance of the proposed method, it was applied on a synthetic signal with network traffic conduct. Firstly, it is required to generate network traffic data. It has been shown that local-area network

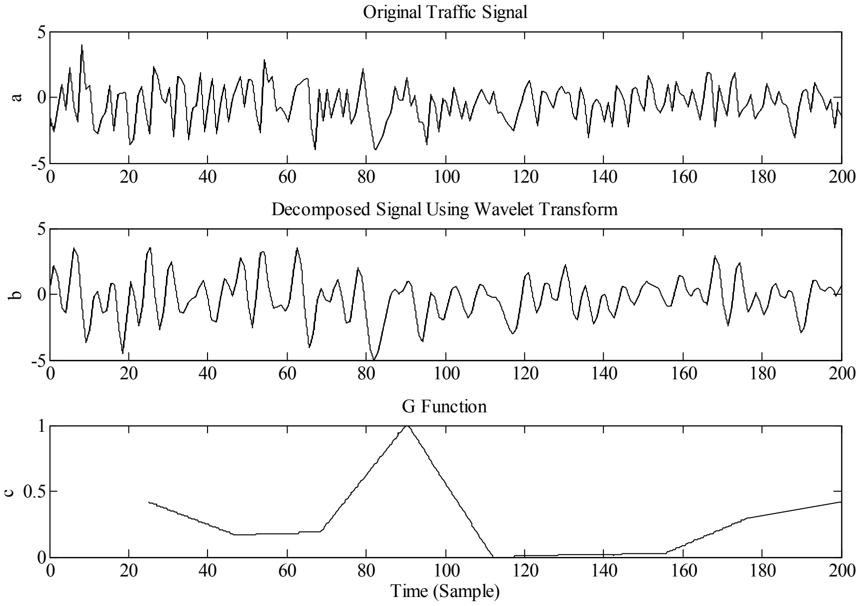

Figure 1. Deviation detection in synthetic signal using proposed approach.

traffic can be modeled using statistically self-similar processes. In [8], an algorithm for generating approximate sample paths for a type of self-similar process known as fractional Gaussian noise (FGN) has been presented. In this algorithm, giving Hurst parameter (H) and number of samples in synthetic data (N), synthetic traffic data can be generated.

In this experiment, H and N are chosen as $H = 0.8$ and $N = 200$, respectively. In order to make deviation in the signal, values of one component is chosen zero after the 100th data point. Therefore, due to the changes in statistical attributes of the synthetic signal, it is expected for this point to be detected by the proposed method. The original signal is initially disintegrated using a one level discrete stationary wavelet transform, and DSWT is performed with Daubechies wavelet of order 1.

In this experiment, using equation (7) leads to achieving an optimum window length with the length of 40 samples. FD of the disintegrated signal and G function are computed using the optimum window length. The threshold value for the proposed method is chosen as $\bar{G} + 2\sigma_G$, where $\bar{G}$ and σ_G represent the mean and standard deviation of G function, respectively. Its value has been experimentally chosen to have a better accuracy in deviation detection. As can be seen from the result, the deviation location has been successfully detected using the proposed method (Figure 1).

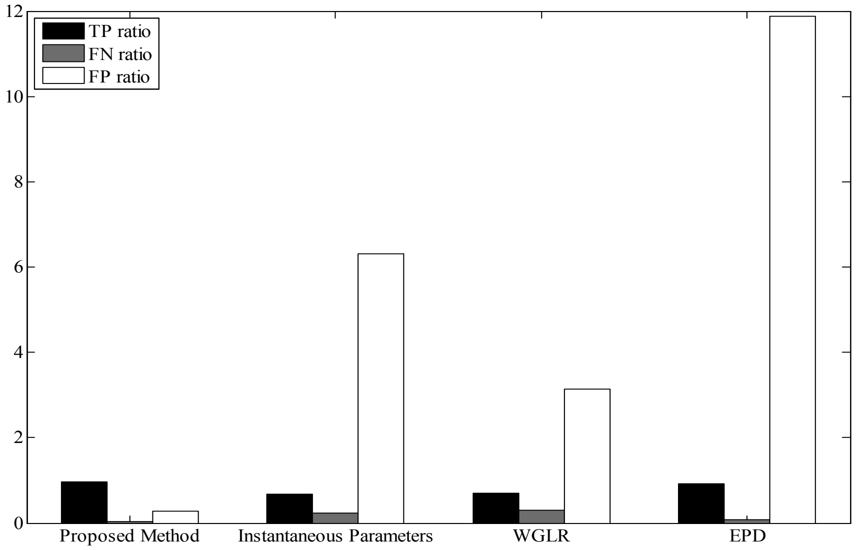

Figure 2. Deviation detection using proposed method in comparison with three other existing methods.

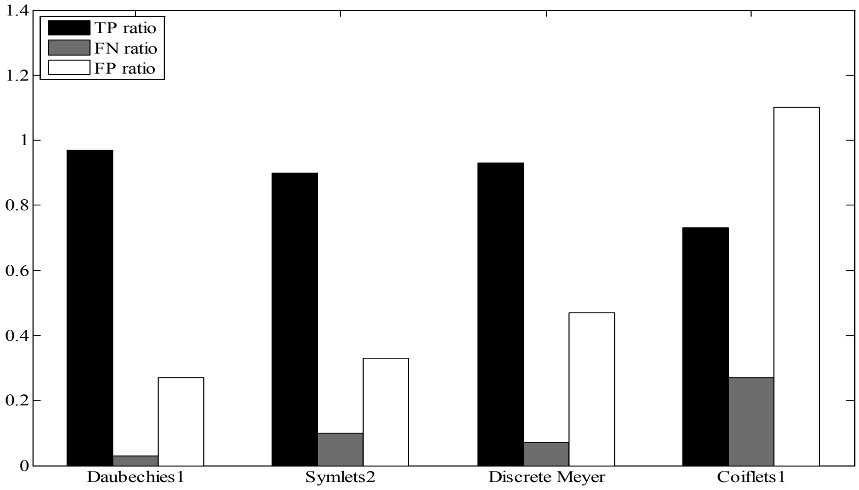

Figure 3. Results of deviation detection using proposed method when four different wavelet basis functions are used.

The existing approaches, the method in [1], WGLR and EPD methods [7], are also applied on the signal. For the purpose of evaluating the performance of the proposed method, the true positive (TP), miss or false negative (FN) and false alarm or false positive (FP) ratios are used as defined below [16]:

$$FP = \frac{N_f}{N},\ FN = \frac{N_m}{N},\ TP = \frac{N_t}{N} \qquad (8)$$

where N_t, N_m, N_f and N represent the number of true, missed, falsely detected and actual number of detected limits, respectively. An efficient deviation detection approach should have a high value for TP ratio and low values for both FN and FP ratios. The proposed method and the three other existing approaches are applied on a set of 200 synthetic signals, but deviation transpires at random position in the signal. In order to compare the performance of different methods, Figure 2 has been presented to address the detecting result belongs to different algorithms. This figure shows that the proposed method has a better accuracy compared to other existing methods. For example, the FP ratio for the proposed method is 26 times lower than the method in [1], 13 times lower than WGLR method [7] and 45 times lower than EPD method [7]. As can be seen from the figure, EPD method offers a high true-positive rate at the expense of a high false positive rate and WGLR method has the lowest TP ratio.

In the following, impact of wavelet basis functions on the deviation detection accuracy is considered. Figure 3 shows the results of applying proposed method on a set of synthetic data when four different wavelet basis functions, namely Daubechies1, Coiflets1, Symlets2 and Discrete Meyer are used. On the basis of experiment we can see that Daubechies1 basis function achieves best results compared to other three wavelets.

5. Conclusions

In this paper a new deviation detection method using wavelet transform and fractal dimension has been introduced. Wavelet transform reveals the important underlying unadulterated form of the data and fractal dimension is a powerful method in transient detection. A comparative study among four typical wavelet basis functions on deviation detection accuracy has been performed. Performance of the proposed method was compared with that of three other existing deviation detection methods. Simulation results showed superiority of the proposed method in deviation detection. The proposed method had the highest value for TP ratio and the lowest value for both FN and FP ratios compared to the existing methods.

References

[1] X. Yao, P. Zhang, J. Gao, G. Hu, Detection of Network Traffic Anomaly Based on Instantaneous Parameters Analysis, *International Conference on Communication Technology, ICCT '06.*, 1–4, (2006).

[2] D. Tran, W. Ma, D. Sharma, Network Anomaly Detection using Fuzzy Gaussian Mixture Models, *International Journal of Future Generation Communication and Networking*, 37–42, (2006).

[3] T. Marina, C. Ji, Adaptive Thresholding for Proactive Network Problem Detection, *IEEE International Workshop on Systems Management*, 5, 108–116, (1998).

[4] V.A. Sotiris, P.W. Tse, M.G. Pecht, Anomaly Detection Through a Bayesian Support Vector Machine, *IEEE Transactions on Reliability*, 59, 277–286, (2010).

[5] J. Luv, X. Li, T. Li, Research on Network Traffic Anomaly Detection Algorithm, *12th IEEE Symposium on Computers and Communications*, 95–100, (2007).

[6] V. Paxson, Fast approximate synthesis of fractional gaussian noise for generating selfsimilar network traffic, *Computer communication review.* 10, 5–18, (1997).

[7] M. Salagean, I. Firoiu, Anomaly Detection of Network Traffic based on Analytical Discrete Wavelet Transform, *8th IEEE International Conference on Communications (COMM)*, 49–52, (2010).

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

[9] L. Lan, L. Gyungho, Ddos attack detection and wavelets, *Telecommunication Systems*, 435–451, (2005).

[10] M.Thangavel, P. Thangaraj, K. Saravanan, Defend versus Anomaly Intrusion Detection Using SWT Mechanism, *International Journal of Innovation, Management and Technology*, Vol. 1, No. 2, 209–213, June (2010).

[11] G. Bachmann, L. Narici, E. Beckenstein, Fourier and Wavelet Analysis, *Springer*, 2002.

[12] G. P. NASON, B. W. SILVERMAN, The stationary wavelet transform and some statistical applications, *Lecture Notes in Statistics*, vol. 103, pp. 281–299, (1995).

[13] J. FALCONER, Fractal Geometry–Mathematical Foundations and Applications, *John Wiley and Sons*, 2003.

[14] M. TYKIERKO, Using invariants to change detection in dynamical system with chaos, *Physica D: Nonlinear Phenomena*, 237, 6–13, (2008).

[15] P. PARAMANATHAN, R. UTHAYAKUMAR, Application of fractal theory in analysis of human electroencephalographic signals, *Computers in Biology and Medicine*, 38, 372–378, (2007).

[16] Y. LI, Y. LE FAN, Q. YE TONG, Endpoint detection in noisy environment using intricacy measure, *Proc. IEEE International Conference*, 3, 1004–1007, (2007).

Received: September, 2011
Revised: March, 2012
Accepted: March, 2012

Contact addresses:

Afshin Shaabany
Science and Research Branch
Islamic Azad University
Fars, Iran
e-mail: afshinshy@yahoo.com

Fatemeh Jamshidi
Science and Research Branch
Islamic Azad University
Fars, Iran
e-mail: fjamshidi59@yahoo.com

AFSHIN SHAABANY was born in Marvdasht, Iran, in 1975. He received the Bs degree in Electerical Engineering from Tehran University, Tehran, Iran in 1999 and the Ms degree in Electrical Engineering from Polytechnic University, Tehran, Iran in 2008. His research interests include IT, telecommunication, switching systems; Intelligent systems.

FATEMEH JAMSHIDI was born in Shiraz, Iran, in 1980. She received the Bs degree in Biomedical Engineering from the Jondi Shapour University, Ahvaz, Iran in 2002 and the Ms and PhD degree in Electrical Engineering from Shiraz University and Tarbiat Modares University, respectively. Her research interests include switching systems; Intelligent systems, Robust control, IT, telecommunication