http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 11 (2): 269-275, 2012
ISSN 1812-5638 / DOI: 10.3923/itj.2012.269.275

Hurst Parameter for Security Evaluation of LAN Traffic

Ruoyu Yan and Yingfeng Wang
College of Computer and Information Engineering, Henan University of Economics and Law Zhengzhou, Henan Province, 450002, China

Abstract: Self-similarity analysis and anomaly detection in networks are research focus in the traffic analysis and network security community. How to use Hurst parameter estimated accurately to detect traffic anomaly timely is a meaningful work. In the study, based on traffic self-similarity, an effective scheme is proposed to detect and evaluate LAN traffic anomaly. Firstly, iterative estimation method is used to estimate Hurst parameters of four traffic metrics in real time. Then the Hurst values are compared with confidence intervals of normal values to detect anomaly in four kinds of traffics. Finally the anomaly detection results are integrated and normalized to evaluate general security situation of LAN traffic. The simulation results show that iterative estimation of Hurst value has faster speed, higher accuracy and smaller confidence intervals than variance-time plots, a widely used time-domain estimation method and than a new whittle estimator method, which is better than wavelet analysis method. On the other hand, experiments on real LAN traffic show that the method proposed in the paper can detect LAN traffic anomaly and display security situation of LAN traffic accurately.

Key words: Anomaly detection, security evaluation, hurst parameter, iterative estimation, self-similarity, LAN traffic

INTRODUCTION

Now-a-days network size expands fast; customers' demand for network business increases rapidly; more and more system vulnerabilities are found and exploited by hackers constantly. All these factors lead to the difficulties of network security management. How to monitor network traffic and find anomaly in real time has important significance of improving on network reliability and availability.

Many previous studies show that very wide range of network traffics, such as traffics in LAN (Leland *et al*., 1993), WAN (Paxson and Floyd, 1995), VBR business (Beran *et al*., 1995), WWW business (Crovella and Bestavos, 1997), exhibit the features of self-similarity. Self-similarity means network traffic displays shape similarity in a very wide time scale. Hurst parameter (abbreviated as H) is an important index to measure the similarity which is often used in traffic congestion control and access control. Therefore, to estimate Hurst parameter accurately and rapidly has significance in network management and control.

The traditional methods for Hurst estimation include: variance-time plots, R/S method, periodogram method and Whittle estimator (Beran, 1994). These methods have their own features and merits. However, by far, wavelet analysis method due to Abry and Veitch (Abry and Veitch, 1998) is an ideal method and widely used. Kettani and Gubner (2002) proposed a new method to compute Hurst value and the experimental results indicate the method is better than wavelet analysis method. Cheng *et al*. (2009) proposed a refined self-similar parameter estimation algorithm to estimate Hurst value which is based on wavelet transform and Hilbert-Huang transform. Then some experiments are done to prove the improvement in the estimation of Hurst value and an example of network traffic anomaly detection is given. The Hurst parameter estimation method is effective but also is time consuming and not quite suitable in real time.

Wang and Fang (2005) analyzed the self-similarity of network traffic data in DARPA99 and found that extreme burst traffic in a specific time period can change self-similarity features on traffic process drastically. This means Hurst parameter can be used to detect burst traffic anomaly. Hong and Jiangang (2009), based on self-similar traffic model have designed a simple anomaly detection engine deployed in Snort Intrusion Detection System. Although they declare that this can improve detection ability effectively, they did not show any experimental results in details.

Hariri *et al*. (2003) analyzes the network performance impact caused by system fault and attack in large scale network system from the perspective of network traffic measurement. Router queue, CPU usage, network

Corresponding Author: Ruoyu Yan, Henan University of Economics and Law, Zhengzhou, Henan Province, 450002, China
Tel: +86-15838318078

bandwidth usage are used as indicators to evaluate whole system's security comprehensively. Hu *et al.* (2005) proposes a method to evaluate data transfer security in network which consider Hurst parameter and the threat weight of network attacks. The method assumes all of the routers have uniform weight obtained only by experience which is not very reasonable. Furthermore, when evaluating security of data transfer path, all routers packets passing by must be considered. These two factors make it difficult to get practical application. Different from Hariri *et al.* (2003) and Hu *et al.* (2005) a security evaluation on LAN traffic is thought to be more feasible and reasonable.

Variance-time plots method and Whittle estimator are used to verify the self-similarity of LAN traffic in Ministry of Education Key Lab for Network Security and it is found that the all traffic, TCP traffic, UDP traffic, ARP traffic have their own conspicuous self-similarity. Based on this and previous analysis, an effective scheme is proposed to detect LAN traffic and evaluate its security. In the scheme, an efficient method called iterative estimation is used to compute Hurst parameter; a Hurst confidence intervals estimation method is used to detect whether Hurst value deviates from normal state or not; based on detection results of four traffic metrics, normalized method is used to evaluate the general security of LAN traffic. All these methods achieve the goals of detecting anomalous traffic and evaluating traffic security in LAN accurately.

LAN TRAFFIC SECURITY EVALUATION SCHEME BASED ON SELF-SIMILARITY

The scheme is composed of five modules shown in Fig. 1: Traffic collection, statistical analysis, Hurst parameter estimation, anomaly detection and normalized security evaluation. Brief description of every module is as follows:

- **Traffic collection:** In order to reduce the impact on normal use of network, when collecting LAN traffic, traffic on router is mirrored to traffic collection server. The traffic collection software running in server captures and processes original traffic. Winpcap development kit which has perfect packet capturing performance is used in the software
- **Statistical analysis:** Packets received from router are unpacked and packet type information is extracted, then four traffic metrics, such as all packet number, TCP packet number, UDP packet number and ARP packet number in LAN are counted, respectively per unit time

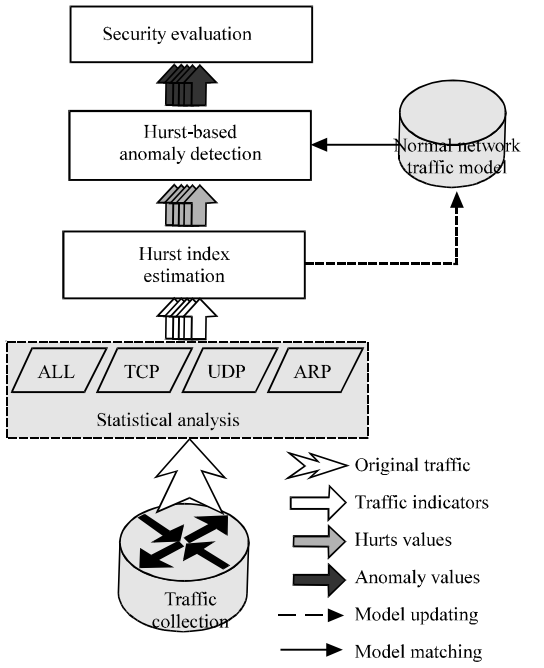

Fig. 1: LAN traffic security evaluation scheme

- **Hurst parameter estimation:** Hurst values of four traffic metrics are calculated by iterative estimation method in real time. The values can be used to detect traffic anomaly and update normal traffic model
- **Anomaly detection:** The current calculated Hurst value is compared to normal traffic model. If the value deviates from normal model, current traffic is thought to be anomalous and is normal otherwise. Normal traffic model is built under the analysis of a period of normal network traffic in LAN monitored. The model includes normal Hurst value and its confidence interval and can be updated during anomaly detection
- **Security evaluation:** Network security risk index is computed by weighted average method which integrates four kinds of traffic anomaly detection results. The risk index reflects the current state of network data transfer security and provides an intuitive security situation for network administrators

HURST PARAMETER ESTIMATION

Definition of self-similar

Definition 1: Let X_n (n = 1,2,3,…) be a discrete stochastic process. If:

$$X_i^{(m)} = \frac{1}{m}\sum_{k=(i-1)m+1}^{im} X_k$$

$X^{(m)}_n$ is called m order aggregated processes of X_n with m order autocorrelation function $\rho^m(k)$.

Definition 2: A wide-sense stationary stochastic X_n (n = 1,2,3,...) is called self-similar if X_n and its aggregated processes of m order $X^{(m)}_n$ have the same autocorrelation function, namely $\rho^m(k) = \rho(k)$ (m = 1,2,3,...). That means $X^{(m)}_n$ and X_n have the same second-order statistical properties.

Autocorrelation function of a wide-sense stationary self-similar process satisfies:

$$\rho_k = H(2H-1)K^{2H-2},\ k\to\infty \quad (1)$$

where, H (0.5<H<1) is Hurst parameter. H is greater, the higher the degree of self-similarity. Because of $\sum_k \rho_k = \infty$, self-similar process is often called long-range correlation. That implies when k is very larger, the time series exists greater relevance. Fractal Gaussian Noise (FGN) process is a typical self-similar process. How to measure Hurst parameter precisely and quickly has very important significance on study of features and changes of traffic self-similar.

Iterative estimation algorithm: If network traffic time series X_i is self-similar, where X_i represents network traffic (measured in packets, bytes etc.) during i-th time period, the autocorrelation function ρ_k satisfies Eq. 1. The iterative formula to calculate H can be obtained after the transformation of Eq. 1:

$$H_{i+1} = \sqrt{(\rho_k k^{2-2H_i} + H_i)\times 0.5},\quad k\to\infty \quad (2)$$

For a given time series $X_1, X_2, \ldots, X_n$, we put:

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and:

$$\hat{\gamma}_k = \frac{1}{n-k}\sum_{i=1}^{n-k}(X_i-\bar{x})(X_{i+k}-\bar{x}),\qquad \hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0},\quad k = 0,1,...$$

to denote the sample mean, sample covariance and sample autocorrelation function of the process X_i. Sample autocorrelation function $\hat{\rho}_k$ is used to replace autocorrelation function ρ_k, then Eq. 2 is rewritten as:

$$\hat{H}_{i+1} = \sqrt{(\hat{\rho}_k k^{2-2\hat{H}_i} + \hat{H}_i)\times 0.5},\quad k\to\infty \quad (3)$$

For long-range correlation process, $\hat{H}_0 = 0.5$ is set. The prerequisite for the establishment of Eq. 3 is that k is infinite. However, experiments show that when k = 1 it not only can obtain Hurst value with enough precision but also can greatly reduce the computation. In addition, it is found that the iteration result is not ideal when k is large. The reason is that with k increasing, the error generated by ρ_k instead of $\hat{\rho}_k$ increases the effect on Hurst estimation. Therefore, k = 1 is set in Eq. 3 and the simplified iteration formula is:

$$\hat{H}_{i+1} = \sqrt{(\hat{\rho}_1 + \hat{H}_i)\times 0.5} \quad (4)$$

Fixed point theorem can prove that Eq. 4 is convergent to unique value in (0.5, 1). The simplified equation is then used to estimate Hurst parameter in the following experiments.

TRAFFIC ANOMALY DETECTION METHOD BASED ON HURST PARAMETER

Under normal circumstances£¬network traffic shows diurnal pattern (that is, network has busy period and little busy period every day) to a certain extent. In order to reduce the impact on Hurst estimation caused by periodicity of network traffic (Akgul *et al*., 2011), it is necessary to process traffic at different time period, respectively. In practice, firstly a week of normal traffic is collected and four traffic metrics mentioned before are extracted. Iterative estimation method is used to compute Hurst values. Then means of a week of normal Hurst values for each metric in each time division in one day are obtained. Next an effective method introduced by Kettani and Gubner (2002) is used to compute 98% confidence intervals of Hurst ($0.5 \le H \le 0.95$). The initial state of normal traffic model is finally established. When starting real time detection, the current calculated Hurst value is compared to normal traffic model for each traffic metric. If the new Hurst value is within the confidence interval of corresponding normal Hurst value, the traffic is thought to be normal and the detection result is 0; if not, anomaly is thought to be found and the detection result is 1. When the detection result is 0, normal Hurst value and its confidence interval are needed to be updated correspondingly in normal traffic model. The updating process is described in Table 1.

Table 1: Normal traffic model updating algorithm

Algorithm: Normal traffic model update

Notations:

curH: current calculated H value

norH: normal Hurst value in normal traffic model

norH_up: the upper limit of 98% confidence interval of norH

norH_low: the lower limit of 98% confidence interval of norH

Initialization:

The initial norH, norH_up and norH_low are computed from a week of normal traffic.

Upon receiving a current calculated H value:

1: get corresponding norH, norH_up and norH_low from normal traffic model;

2: if norH_low≤curH≤norH_up then

3: update norH to 0.5 norH + 0.5curH;

4: compute 98% confidence interval of norH;

5: update norH_up and norH_low of norH;

6: end if

NORMALIZED SECURITY EVALUATION METHOD

Although there are various forms of attacks in LAN, almost all these attacks are implemented on TCP, UDP or ARP protocols. That is why TCP, UDP and ARP packets are taken into consideration for anomaly detection. Of course to detect the all packets in LAN is also important and necessary. Hence four traffic metrics are selected as observations and the set of four indicators are obs = {1-ALL packets, 2-TCP packets, 3-UDP packets, 4-ARP packets}. Weighted average method is used to integrate the detection results of four traffic metrics and the network security risk index $F_{traffic}$ can be calculated as follows:

$$F_{traffic} = \sum_{i=1}^{4} w(i)\, F_{obs}(i), \quad \sum_{i=1}^{4} w(i) = 1 \tag{5}$$

where, w (i) is weight of obs (i). Generally anomaly in all traffic is paid more attention to and other three types of traffic anomalies are treated equally. Thus weight vector is set to w = [0.4, 0.2, 0.2, 0.2]. F_{obs} (i) is the detection result of traffic metric obs (i). If at current period the Hurst value of traffic metric obs (i) is within the confidence interval of normal Hurst, we have F_{obs} (i) = 0 and F_{obs} (i) = 1 otherwise. The network security risk index $F_{traffic}$ is the final indicator to judge network traffic normal or not if threshold is set. When risk index exceeds threshold, alarm can be raised.

THE COMPARISON TO THREE KINDS OF HURST PARAMETER ESTIMATION METHODS

Matlab is used to generate 100 realizations of a FGN for each value of H = 0.5, 0.6, 0.7, 0.8, 0.85, 0.9, 0.95. The length of each realization is n = 5000 points. Iterative estimation method is compared with variance-time plots and method mentioned by Kettani and Gubner (2002) (the method is called Houssain method in the following). Houssain method is better than wavelet analysis method, a widely used time-domain and frequency-domain estimation method (Kettani and Gubner, 2002). For iterative estimation method the iteration end condition is set to H_{i+1}-$H_i \le 0.0005$ and generally Hurst value can be obtained after six iterations. For a given estimation method, 100 estimated values of H are obtained. Their sample mean $\hat{H}$, sample variance are computed and the 95% confidence intervals of the estimate $\hat{H}$ are also provided. The comparison result of the application of three methods to these data sets is given in Table 2.

Table 2: Results of three methods using 100 independent realizations with length of 5000

	Variance-time plots		Houssain method		Iterative estimation	
H	$\hat{H}$	Conf. Inter.	$\hat{H}$	Conf. Inter.	$\hat{H}$	Conf. Inter.
0.5	0.493	[0.455,0.531]	0.499	[0.472,0.526]	0.502	[0.471,0.533]
0.6	0.577	[0.542,0.612]	0.610	[0.585,0.635]	0.615	[0.595,0.635]
0.7	0.684	[0.643,0.725]	0.712	[0.684,0.740]	0.711	[0.693,0.729]
0.8	0.783	[0.739,0.827]	0.792	[0.771,0.813]	0.805	[0.787,0.823]
0.85	0.817	[0.772,0.862]	0.841	[0.816,0.866]	0.853	[0.837,0.876]
0.9	0.854	[0.807,0.901]	0.873	[0.846,0.900]	0.884	[0.863,0.905]
0.95	0.892	[0.845,0.939]	0.910	[0.883,0.937]	0.921	[0.899,0.943]

Table 3: Results of three methods using 100 independent realizations with length of 500 and length of 1000

	H					
	0.5		0.7		0.9	
Data length	500	1000	500	1000	500	1000
$\hat{H}$ (Variance-time plots)	0.484	0.490	0.662	0.674	0.796	0.819
$\hat{H}$ (Houssain method)	0.486	0.494	0.685	0.689	0.845	0.854
$\hat{H}$ (Iterative estimation)	0.487	0.497	0.704	0.706	0.859	0.866

From Table 2, it is observed that with the increasing of H value, iterative estimation method is more accurate to estimate H value than other two methods and confidence intervals obtained by the method are narrower than those obtained by other two methods. This means iterative estimation method has much stability and accuracy. In the experiment, variance-time plots, Houssain method and iterative estimation method spend 2.160, 0.881 and 0.002 sec to estimate a Hurst value in average, respectively. Clearly iteration estimation method is much faster than other two methods and is better to estimate Hurst parameter in real time.

In addition, 100 realizations of a FGN with data length of 500 and 1000, respectively are generated for each value of H = 0.5, 0.7, 0.9. Again for a given estimation method, for each data length, 100 estimated values of H are obtained and their sample mean $\hat{H}$ is computed. The comparison result of the application of three methods to these data sets is given in Table 3. From Table 3, it is observed that although the sample data has short length, estimate of iterative estimation method is closer to the true

H value than other two methods. This indicates iterative estimation method is more suitable for short length data which reduces computation.

In short, iterative estimation method is full of potential and we use it to estimate Hurst parameter in practice.

ANOMALY DETECTION AND SECURITY EVALUATION OF LAN TRAFFIC

Test environment: The experimental network set up in Ministry of Education Key Lab for Network Security is shown in Fig. 2.

In the network topology: the device with IP address 202.117.14.242 is a router; the server with IP address 202.117.14.243 provides web services to external users; the server with IP address 202.117.14.244 provides file transfer service and mail service; other computers provide general applications to internal users. The attack program is installed in host with IP address 192.168.2.33. Traffic collection and analysis program is installed in host with IP address 192.168.2.55.

According to Leland *et al.* (1993), self-similarity of a week of all packets, TCP packets, UDP packets and ARP packets sampled at 10, 1, 0.1 sec, respectively is analyzed in test network and it is found that traffic sampled at 0.1 sec reflects the more obvious self-similarity. Upon that 0.1 sec are selected as traffic statistics time bin in practice and Hurst value is computed once every 1 min. The means of Hurst values of four traffic metrics are shown in Table 4. It is observed that ALL traffic and TCP traffic has very similar Hurst value, because TCP traffic is main component of ALL traffic. ARP traffic has smaller Hurst value because of relatively sparse ARP packets in LAN.

Analysis and discussion of experimental results: In order to test and verify the effectiveness of the method proposed in the paper, a SYN-flooding attack against WWW server with IP address 202.117.14.243 is launched which lasts for one minute, then 10 min later a trinoo attack against FTP server with IP address 202.117.14.244 is launched which lasts for two minutes. SYN-flooding attack sends TCP packets to the attacked host and trinoo attack sends UDP packets to the attacked host. They are all Denial of Service (DoS) attacks which can lead to the collapse of the attacked host systems.

During the experiment, Hurst parameters of four traffic metrics are estimated every 1 min. The detection and security evaluation results are shown in Fig. 3-5. In Fig. 3 and 4, the red solid lines represent upper and lower limits of 98% confidence intervals of normal Hurst values ranging from 0.5 to 0.95. Each traffic metric has a marker with coordinates in Fig. 3 and 4, among which vertical ordinate value represents currently estimated value of Hurst and horizontal ordinate value represents its corresponding normal value of Hurst. Figure 3 shows deviations of H values of four metrics from normal range in the TCP-flooding attack. The Hurst values of ALL metric and TCP metric deviate from the normal confidence intervals very far because of changes in TCP traffic characteristics caused by SYN-flooding attack. Accordingly in Fig. 5a the anomaly value is one in ALL packets metric and TCP packets metric situation curves

Table 4: The means of Hurst values of four traffic metrics

Traffic metric	Hurst value
ALL	0.831
TCP	0.824
UDP	0.667
ARP	0.563

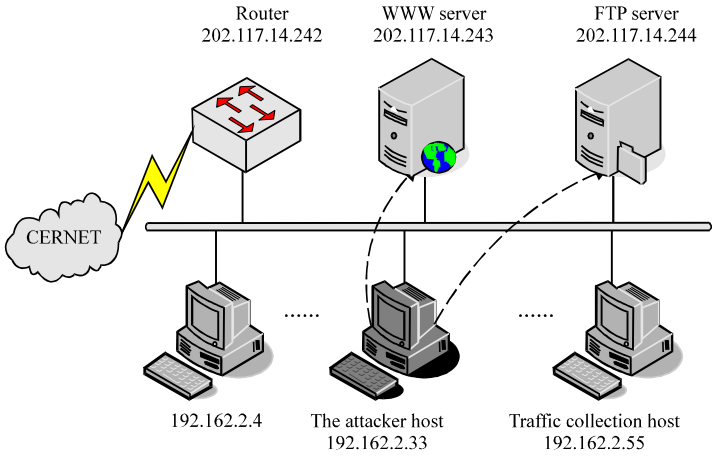

Fig. 2: Test network topology

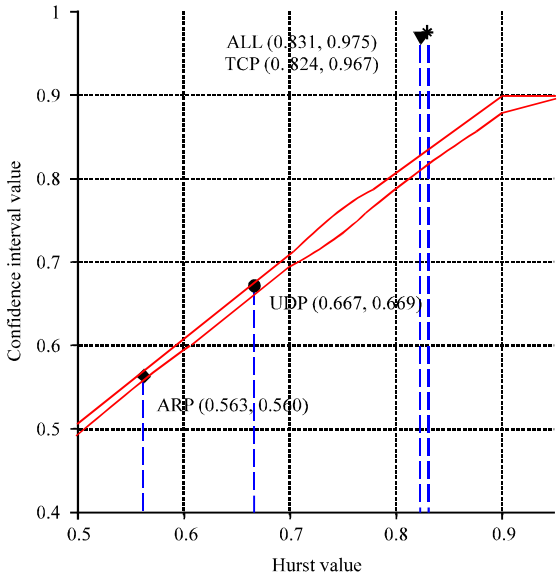

Fig. 3: Deviations of H values in the TCP-flooding attack

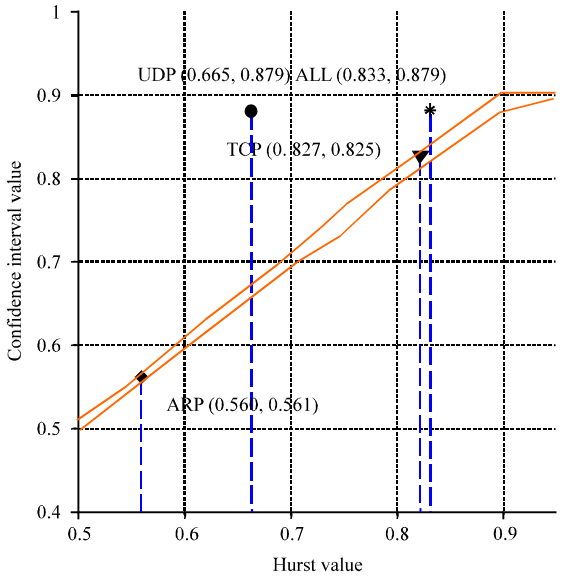

Fig. 4: Deviations of H values in the trinoo attack

from 1 to 2 time bin. Similarly, Fig. 4 shows deviations of H values of four metrics from normal range in the trinoo attack. The Hurst values of ALL metric and UDP metric deviate from the normal confidence intervals very far because of changes in UDP traffic characteristics caused by trinoo attack. Accordingly in Fig. 5a the anomaly value is one in ALL packets metric and UDP packets metric situation curves from 12 to 14 time bin.

It is thought that in normal network operations, for a particular node, data received from a large number of

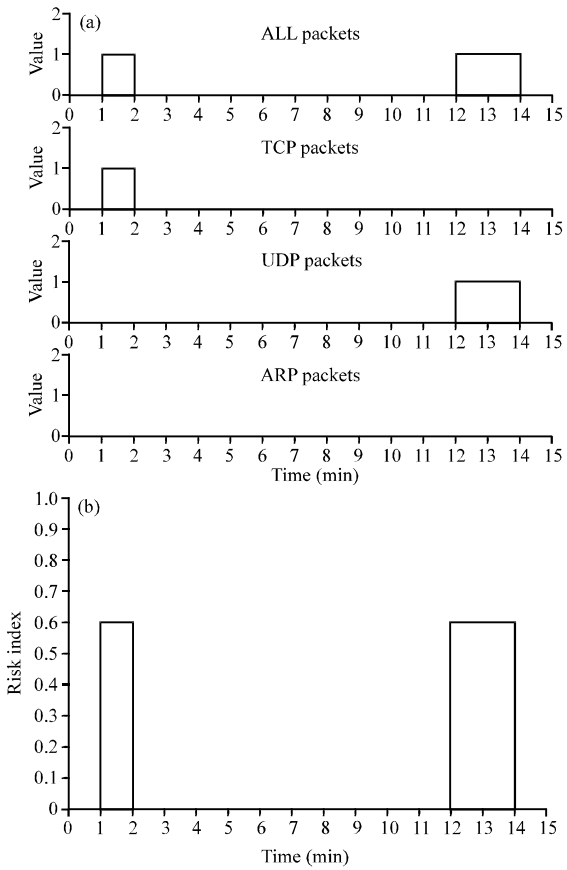

Fig. 5(a-b): Security situation of network traffic in LAN; (a) Security situation of four kinds of network traffic and (b) General security situation of network traffic

hosts don't have time correlation and packets characteristics (such as protocol types, packet size) correlation. But according to DoS/DDoS attack principle, a large number of data packets requests coming from attack agents controlled by hacker will lead to time correlation and packets characteristics correlation. And these correlations give rise to significant change in Hurst parameter. Thus the attacks can be detected by method proposed in the study.

Figure 5 shows traffic anomaly detection situation over a period of time. Among it, Fig. 5a shows anomaly detection results of four traffic metrics which intuitively reveals what types of traffics are affected by potential attacks. Figure 5b shows general security risk index curves which intuitively reveals the risk level of network at different time, the higher the index value, the higher the risk of the network being attacked. In short, Fig. 5 provides an intuitive security risk report about LAN traffic for network system administrators.

This helps us to identify the security factors and take appropriate preventive measures.

CONCLUSIONS

This study proposes a scheme based on self-similarity in LAN traffic to detect anomaly and evaluate security of traffic. Two main contributions are as follows:

- Propose using iterative estimation method to estimate Hurst parameter and simulation experiments demonstrate that the method is faster, more accurate and more suitable for short length data than other two typical estimation methods
- Propose a new method to detect LAN traffic anomaly, in which confidence interval of normal Hurst value is computed to judge traffic abnormal or not and DoS/DDoS attack detection experiments demonstrate that the method can detect anomaly and display security situation of LAN traffic accurately

For the purpose of improving in security evaluation performance, in the future some experiments should be done to find a method to adjust weigh vector of four traffic metrics in real time; in addition, when anomaly is detected, how to pinpoint abnormal traffic responsible for the anomaly in LAN is also future work.

ACKNOWLEDGMENTS

This study is supported mainly by Henan Science and Technology Research Projects of China (112102210393) and Henan Fundamental and Advanced Technology Research Projects of China (112300410201).

REFERENCES

Abry, P. and D. Veitch, 1998. Wavelet analysis of long-range-dependent traffic. IEEE Trans. Inform Theory, 44: 2-15.

Akgul, T., S. Baykut, M. Erol-Kantarci and S.F. Oktug, 2011. Periodicity-based anomalies in self-similar network traffic flow measurements. IEEE Trans. Instrument. Measure., 60: 1358-1366.

Beran, J., 1994. Statistics for Long-Memory Processes. Chapman and Hall, New York.

Beran, J., R. Sherman, M.S. Taqqu and W. Willinger, 1995. Long-range dependence in variable-bit-rate video traffic. IEEE Trans. Commun., 43: 1566-1579.

Cheng, X., K. Xie and D. Wang, 2009. Network traffic anomaly detection based on self-similarity using HHT and wavelet transform. Proceedings of the 5th International Conference on Information Assurance and Security, August 18-20, 2009, Xian, China, pp: 710-713.

Crovella, M. and A. Bestavos, 1997. Self similarity in world wide web traffic: Evidence and possible causes. IEEE/ACM Trans. Network., 5: 835-846.

Hariri, S., Q. Guangzhi, T. Dharmagadda, M. Ramkishore and C.S. Raghavendra, 2003. Impact analysis of faults and attacks in large-scale networks. IEEE Security Privacy, 1: 49-54.

Hong, K. and Z. Jiangang, 2009. An improved snort intrusion detection system based on self-similar traffic mode. Proceedings of the International Symposium on Computer Network and Multimedia Technology, January 18-20, 2009, Wuhan, pp: 1-4.

Hu, H., W. Guo, B. Zhang and X. Chen, 2005. A method of security measurement of the network data transmission. Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium, April 4-8, 2005, Denver, Colorado, pp: 8.

Kettani, H. and J.A. Gubner, 2002. A novel approach to the estimation of the Hurst parameter in self-similar traffic. Proceedings of the 27th Annual IEEE Conference on Local Computer Networks, November 6-8, 2002, Tampa, USA., pp: 160-165.

Leland, W., M. Taqqu, W. Willinger and D. Wilson, 1993. On the self-similar nature of Ethernet traffic. Proceedings of the ACM SIGCOMM'93 Conference on Communications Architectures, Protocols and Applications, September 13-17, 1993, San Francisco, CA.

Paxson, V. and S. Floyd, 1995. Wide area traffic: The failure of Poisson modeling. IEEE/ACM Trans. Network., 3: 226-244.

Wang, X. and B.X. Fang, 2005. An exploratory development on the Hurst parameter variety of network traffic abnormity signal. J. Harbin Inst. Technol., 37: 1046-1049.