LoSS Detection Approach Based on ESOSS and ASOSS Models

Mohd Fo'ad Rohani[1], Mohd Aizaini Maarof[2], Ali Selamat[3] and Houssain Kettani[4]
[1,2,3]*Faculty of Computer Science and Information Systems*
Universiti Teknologi Malaysia
81300 Skudai, Johor, Malaysia
[4]*Department of Electrical and Computer Engineering and Computer Science*
Polytechnic University of Puerto Rico
P. O. Box 192017
San Juan, Puerto Rico 00919, USA
Email: {foad, aizaini, aselamat}@utm.my, hkettani@pupr.edu

Abstract

This paper investigates Loss of Self-similarity (LoSS) detection performance using Exact and Asymptotic Second Order Self-Similarity (ESOSS and ASOSS) models. Previous works on LoSS detection have used ESOSS model with fixed sampling that we believe is insufficient to reveal LoSS detection efficiently. In this work, we study two variables known as sampling level and correlation lag in order to improve LoSS detection accuracy. This is important when ESOSS and ASOSS models are considered concurrently in the self-similarity parameter estimation method. We used the Optimization Method (OM) to estimate the self-similarity parameter value since it was proven faster and more accurate compared to known methods in the literature. Our simulation results show that normal traffic behavior is not influenced by the sampling parameter. For abnormal traffic, however, LoSS detection accuracy is very much affected by the value of sampling level and correlation lag used in the estimation.

1. Introduction

Internet traffic monitoring especially to detect anomaly traffic behavior presence in the network is essential to ensure uninterrupted Internet services. There are several anomaly traffic detection models that have been applied in the previous works including statistical moments (i.e. mean and standard deviation models), multivariate model and time series model [13]. When dealing with huge amount of traffic packets where Internet behavior keeps changing, anomaly traffic detection using time series model is preferred since the model can produce better results than others statistical models [13]. Recent studies have shown that self-similarity model is widely used for Internet traffic modeling and analysis [3], [4] [7], [9], [11], [12]. According to self-similarity model, the autocorrelation of inter arrival traffic packets is assuming to exhibit hyperbolic decay and Long Range Dependent (LRD). These assumptions are true for normal traffic but in the presence of malicious traffic such as Denial of Service (DoS) packets, the self-similarity distribution error [15] is introduced. Consequently, Loss of Self-similarity (LoSS) is detected [1], [6], [15] and this can be used to alert security analysts the presence of uncontrolled self-similarity structure in the network [4]. As a result, packets queue buffer delay and packets drop rates will drastically increased [4] hence degrading Quality of Service (QoS) performance.

Implementation of LoSS detection with Second Order Self-Similarity (SOSS) statistical model has been introduced due to high speed and accuracy needs [6]. Previous work [6], however, used fixed sampling time series packets which is insufficient to reveal self-similarity distribution error correctly [14]. In this work, we investigate LoSS detection accuracy and its dependency on two variables known as sampling (or aggregation) level and correlation lag. This is important when combining the idea of exact and asymptotic self-similarity models concurrently in the parameter estimation method. In exact and asymptotic self-similarity models, sampling level does not affect the self-similarity parameter estimation for normal traffic. In the presence of self-similarity distribution error such as exists in malicious traffic traces, however, their autocovariance and variance decay are not identical. Thus, it can be shown that sampling level

978-0-7695-3324-7/08 $25.00 © 2008 IEEE
DOI 10.1109/IAS.2008.37

Proceedings of the 4th International Conference on Information Assurance and Security (IAS 2008), pp. 192-197, Naples, Italy, September 2008.

and correlation lag are two important parameters that can reveal LoSS detection accurately.

In this work, we use Second Order Self-similarity (SOSS) statistical model and the Optimization Method (OM) that was developed in [7] to estimate the self-similarity parameter or Hurst (*H*) value. According to the previous work [6], LoSS is detected if the Curve Fitting Error (CFE) estimated using OM exceeds the threshold value. We presume anomaly traffic detection based on LoSS and CFE criterion will suffer high false alarm detection rate if improper sampling level and insufficient correlation lag process are used. Thus, this work is aimed to analyze LoSS detection performance using different values of sampling level and correlation lag. The effect of these parameters in minimizing the CFE accuracy and estimation time speed is investigated. This paper is organized as the following. Section 2 discuses in brief the mathematical definitions of self-similarity models and LoSS detection method. The experimental procedure and empirical analyses are presented in Section 3. Finally, the conclusions are summarized in Section 4.

2. Self-Similarity Model and Estimation Method

2.1 SOSS Model and Estimation Method

Let us define a second-order stationary process $X = \{X(t), t > 0\}$ with constant mean μ, finite variance σ^2 and autocorrelation $\rho(k)$ as follow:

$$\mu = E[X(t)], \quad \sigma^2 = E[(X(t) - \mu)]^2 \tag{1}$$

$$\rho(k) = E[(X(t) - \mu)(X(t+k) - \mu)] / \sigma^2 \tag{2}$$

Let $X^{(m)} = \{X^{(m)}(t), t > 0\}$ denotes the aggregation process of *X(t)* at aggregation level $m > 0$. Thus, we have:

$$X^{(m)}(t) = \frac{1}{m} \sum_{w=m(t-1)+1}^{mt} X(w) \ , t > 0 \tag{3}$$

Let $\rho^{(m)}(k)$ denotes the autocorrelation function of $X^{(m)}$. X is called Exact Second-Order Self-Similar (ESOSS) if $\rho(k) = \rho^{(m)}(k)$ for all $m \geq 1$. In ESOSS model, the autocorrelation structure is preserved for all *m* such that:

$$\rho(k) = \frac{1}{2}[(k+1)^{2-\beta} - 2k^{2-\beta} + (k-1)^{2-\beta}] \tag{4}$$

where $k > 0$ and $0 < \beta < 1$. X is called Asymptotical Second-Order Self-Similar (ASOSS) if

$$\lim_{m,k\to\infty} \rho^m(k) \sim \frac{1}{2}[(k+1)^{2-\beta} - 2k^{2-\beta} + (k-1)^{2-\beta}] \tag{5}$$

where $k > 0$, $m > 0$ and $0 < \beta < 1$. X is called Long-Range Dependent (LRD) if its autocorrelation function satisfies: $\rho(k) \sim ck^{-\beta}$ where $k \to \infty$, $c > 0$ and $0 < \beta < 1$.

There are several methods to estimate *H*. In this paper we use OM that was developed in [7] which was proven relatively fast and accurate compared to other methods such as the wavelet method. The OM defines Curve-Fitting Error (CFE) function as $E_K(\beta)$ such as:

$$E_K(\beta) = \frac{1}{4K} \sum_{k=1}^{K} (\rho(k) - \rho_n(k))^2 \tag{6}$$

where $\rho(k)$ denotes the autocorrelation function of the model with parameter β that we would like to fit the data to, $\rho_n(k)$ is the sample autocorrelation function of the data, and *K* is the largest value of *k* such that it minimize the edge effect for the calculation of $\rho_n(k)$. If the minimum of $E_K(\beta)$ is less than 10^{-3}, then the data fits the model and the minimizer $\hat{\beta}$ is picked to be the estimate of the parameter β [7].

2.2 LoSS Detection Using Parameter's Adjustment Based on ASOSS Model

Let *X(t)* be a stochastic time series processes with second order stationary that follow self-similarity features. The autocovariance decays of *X(t)* and aggregated $X^{(m)}(t)$ should follow ESOSS model as shown in equation (7):

$$\lim_{m,k\to\infty} \gamma^m(k) = \gamma(k) \sim C_0 k^{-\beta} \tag{7}$$

where *m* is sampling level, *k* is correlation lag, C_o is constant and β is self-similarity parameter. In real Internet traffic packets, however, the self-similarity processes is also considered as processes *x(j)* in the class *X* of those stationary processes that feature as asymptotic decays in autocovariance [10]. Thus, we should consider ESOSS and ASOSS models concurrently in order to estimate the self-similarity parameter correctly for normal and abnormal traffic.

Equations (8), (9) and (10) define autocovariance, variance and autocorrelation for aggregated process $X^{(m)}(t)$ [10] as follow:

$$\lim_{m,k\to\infty} \gamma^m(k) \sim C_1 m^{-\beta} k^{-\beta} \tag{8}$$

$$\lim_{m\to\infty} \gamma^m(0) \sim C_2 m^{-\beta} \tag{9}$$

Proceedings of the 4th International Conference on Information Assurance and Security (IAS 2008), pp. 192-197, Naples, Italy, September 2008.

$$\lim_{m,k\to\infty} \rho^m(k) = \lim_{m,k\to\infty} \left(\frac{\gamma^m(k)}{\gamma^m(0)} \right)$$

$$\sim \frac{C_1 m^{-\beta} k^{-\beta}}{C_2 m^{-\beta}} \sim C_3 k^{-\beta} \qquad (10)$$

where C_1, C_2 and C_3 are constants.

The self-similarity parameter estimation from previous works [7], [9], [11] have used assumption that normal Internet traffic follows ESOSS model which its characteristics should follow equations (8) to (10). Equation (10) clearly demonstrates that autocorrelation decay does not affected by aggregation level (m) parameter. On the other hand, correlation lag (k) plays an important role to obtain high accuracy of self-similarity parameter (β) estimation. In theory, k equal to one is enough to estimate β provided the self-similarity behavior pattern is known prior to the estimation process. In practice, however, real Internet traffic behaviors are mixed from many sources such as various Internet applications, user's actions, network infrastructure [3] and traffic models [12]. Furthermore, modern Internet traffic can go beyond normal self-similarity or LRD behavior such as multi-fractal and chaotic behavior [2], [5]. Therefore, suitable value of k is needed in order to accurately estimate CFE without sacrificing negligible long tail correlation value.

In the presence of malicious traffic such as DoS packets, high intensity of DoS packets can disturb Internet traffic behavior and produce self-similarity distribution error. Thus, normal characteristics of equations (8) to (10) are not valid where LoSS can be detected. Equation (11) shows that for abnormal traffic, the autocovariance and variance decaying pattern of $C_1 m^{-\beta}$ and $(C_2 m^{-\beta})'$ are not identical hence not following the normal self-similarity pattern as shown in equation (10).

$$\lim_{m,k\to\infty} \rho^m(k) = \lim_{m,k\to\infty} \left(\frac{\gamma^m(k)}{\gamma^m(0)} \right) \sim \left(\frac{C_1 m^{-\beta}}{(C_2 m^{-\beta})'} \right) k^{-\beta}$$

$$\neq C_3 k^{-\beta} \qquad (11)$$

This shows that for detecting malicious traffic, aggregation and correlation lag are two parameters that need to be considered for estimating the CFE value correctly in order to improve LoSS detection accuracy. We investigate parameter adjustment that considers sampling level at micro sampling range (i.e. below one second) [2], [5] which known as engineering factor [2] for Internet protocol design. Meanwhile, we limit the change value of correlation lag below one thousand in order to avoid longer time estimation and maintain high speed LoSS detection performance.

3. Experimental and Empirical Analyses

3.1Experimental Datasets

We have evaluated the proposed LoSS analysis approach with different datasets which includes synthetic fractional Gaussion noise (fGn) [8], University of North Carolina (UNC) 2002 and 2003 datasets [11], and Internet traffic FSKSMNet 2006 dataset [14]. The synthetic fGn trace is an artificial traffic packets that exhibit ESOSS model and about 180,000 packets are generated for the testing. On the other hand, UNC2002 is a suspicious trace that was captured in the UNC network infrastructure while UNC2003 represents a normal traffic trace. We have simulated the FSKSMNet traffic traces as to compare with current Internet traffic packets in the Faculty of Computer Science and Information Systems (FCSIS). Our traffic simulation is about 30 minute's duration which consists of normal traffic that was filtered by network firewall and administration policy of FCSIS. On the other hand, our malicious simulation traffic contains TCP SYN and UDP flood attacks at controlled rates. The percentage protocol for normal traffic presented as almost 97% is dominated by TCP protocol while UDP is less than 2.5%. The ICMP, IGMP and others protocol are less than 0.5%. The malicious traffic, on the other hand, contains protocol of 28% TCP SYN and 27% UDP flooding while normal protocols of TCP and UDP are 41% and 3%. The remainder are ICMP, IGMP and others which less than 0.5%. We sample all the simulation traces with sampling level 10ms < m < 1000ms [14]. Meanwhile, we use the correlation lag parameter with the value changes from 50 up to 500. In the process of LoSS detection purposes, we use the criterion of normal behavior as CFE < 10^{-3} [7] while abnormal behavior as CFE > 10^{-3} where LoSS is detected [6].

3.2 LoSS Detection for Synthetic fGn

Figure 1 shows the result of LoSS estimation for fGn trace with different sampling level (m) and correlation lag (k) values. The result clearly shown that zero LoSS is detected for the synthetic trace since the CFE estimation is always below the threshold regardless of m and k values are used. This shows that the fGn trace follows ESOSS model where sampling parameter m does not affect the estimation process of self-similarity parameter. Thus, the traffic behavior estimated using small or high m will provide similar behavior as shown by equation (4). Equation (4) shows parameter that can influence CFE accuracy is correlation lag k. Our result, however, demonstrates the

Proceedings of the 4th International Conference on Information Assurance and Security (IAS 2008), pp. 192-197, Naples, Italy, September 2008

CFE accuracy for fGn trace is not exceeding threshold value despite using various incremented values of *k*.

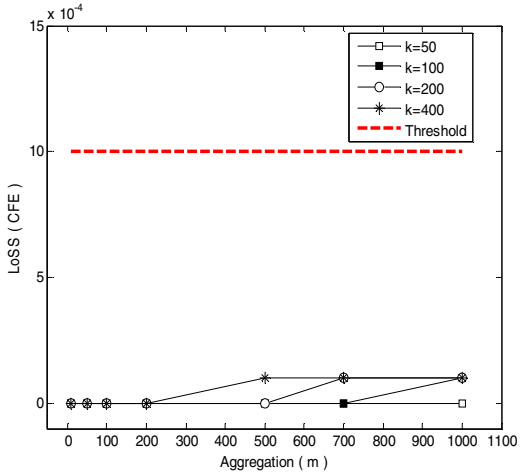
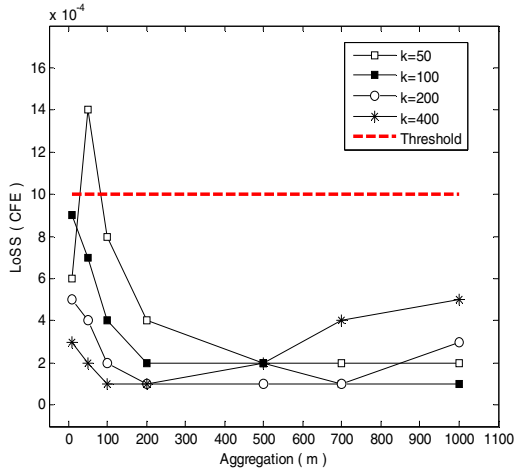


Figure 1. LoSS detection performance for fGn

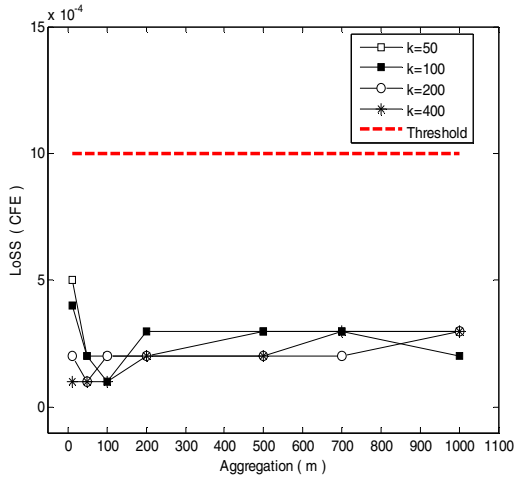
3.3 LoSS Detection for Normal Traffic

The results of LoSS detection for normal Internet traffic simulation traces are shown in Figure 2 (a) for UNC2003 and (b) for FSKSMNet 2006 respectively. The details of CFE estimation for UNC2003 trace are presented in Table 1. It is clearly illustrated in Figure 2(a) that normal of UNC2003 trace is almost following ESOSS model as shown in equation (4). The CFE estimation at all *m* and *k* is estimated below the threshold value except for *m* less than 100ms. As shown in Table 1, the self-similarity parameter of normal trace is incorrectly estimated if small sampling level such as *m* equal 50ms and small correlation lag such as *k* equal 50 are used. Meanwhile, Figure 2(b) shows that the normal FSKSMNet traffic trace follows ESOSS model where their CFE estimation is below the threshold for all *m* and *k* values. The results demonstrate that the estimated CFE values do not exceed the threshold even though *k* is increased to higher value such as 500 for all *m*. The increment of *k* at each value of *m*, however, increases the accuracy of CFE estimation towards fitting the SOSS model.

Our results demonstrate that for normal Internet traffic that follows ESOSS model, *m* does not influence the self-similarity parameter estimation accuracy. This characteristic is clearly shown by equations (4) and (10) to represent normal traffic behavior. The results also agree with previous works in [7], [9] that fixed sampling can be used to estimate the self-similarity parameter correctly, provided the normal traffic trace follows ESOSS model. Thus, it is sufficient for normal trace to apply sampling level such as 10ms or 100ms [9], [11] together with small correlation lag such as 50 [7] to estimate the self-similarity parameter correctly.



(a)



(b)

Figure 2. LoSS detection performance for normal traffic traces: (a) UNC2003 (b) FSKSMNet 2006

Table 1. Details CFE estimation for normal: UNC2003

Sampling (m)	Auto-correlation lag (k)					
	50	100	200	300	400	500
10	0.0006	0.0009	0.0005	0.0004	0.0003	0.0002
50	0.0014	0.0007	0.0004	0.0002	0.0002	0.0002
100	0.0008	0.0004	0.0002	0.0002	0.0001	0.0001
200	0.0004	0.0002	0.0001	0.0001	0.0001	0.0001
500	0.0002	0.0002	0.0001	0.0001	0.0002	0.0002
700	0.0002	0.0001	0.0001	0.0002	0.0004	0.0004
1000	0.0002	0.0001	0.0003	0.0005	0.0005	0.0004

Proceedings of the 4th International Conference on Information Assurance and Security (IAS 2008), pp. 192-197, Naples, Italy, September 2008

3.4 LoSS Detection for Abnormal Traffic

Parameters adjustment which involves m and k in equation (11) can significantly influence LoSS detection accuracy. The result of LoSS detection pattern for suspicious trace UNC2002 is illustrated in Figure 3 while the details are presented in Table 2. It is clearly shown in Figure 3 that different combination of m and k used in estimation method can either detect LoSS successfully or LoSS is undetected. The results show that at each level of m, different values of k can give different estimation of CFE accuracy. Thus, improper values of m and k chosen in the estimation method can contribute high false alarm LoSS detection.

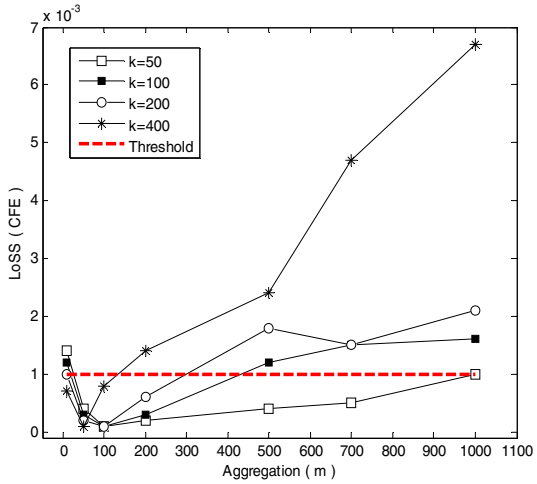


Figure 3. LoSS detection for suspicious traffic UNC2002

Table 2. Details CFE estimation for suspicious: UNC2002

Sampling (m)	Auto-correlation lag (k)					
	50	100	200	300	400	500
10	0.0014	0.0012	0.001	0.0008	0.0007	0.0007
50	0.0004	0.0003	0.0002	0.0002	0.0001	0.0001
100	0.0001	0.0001	0.0001	0.0004	0.0008	0.0007
200	0.0002	0.0003	0.0006	0.0012	0.0014	0.0016
500	0.0004	0.0012	0.0018	0.0016	0.0024	0.0035
700	0.0005	0.0015	0.0015	0.0024	0.0047	0.0063
1000	0.001	0.0016	0.0021	0.0055	0.0067	0.0071

As shown in Table 2, at small m such as 10ms and k uses less than 100, LoSS is detected for the suspicious trace. However, the traffic trace follows SOSS model if k bigger than 100 is used. Meanwhile, the trace has zero LoSS detection at m equal 50ms and 100ms when k less than 500 is used. Thus, our result demonstrates that the LoSS occurrence can be revealed efficiently if we increase sampling level to higher value such as larger than 200ms. This can be shown in Table 2 such as at sampling level m equal 200ms and k bigger than 200, LoSS is fully detected. Similarly, the LoSS occurrence is also entirely revealed and detected at m larger than 500ms and k bigger than 50 are used.

On the other hand, the results of LoSS detection for FSKSMNet 2006 trace is illustrated in Figure 4 and the details are presented in Table 3. Figure 4 demonstrates similar pattern with Figure 3 where LoSS detection accuracy is very much depending on parameter m and k. As shown in Table 3, the malicious trace follows normal SOSS model at small sampling level such as m equal 10ms for all values of k. The increment of sampling level, however, can improve the accuracy of LoSS detection as in the previous results of UNC2002 trace. For instance, at m equal 50ms, 100ms and 200ms; LoSS is revealed when using k bigger than 300, 100 and 50 respectively. Meanwhile, at m larger than 500ms LoSS is fully detected regardless any values of k between $50 < k < 500$ are used.

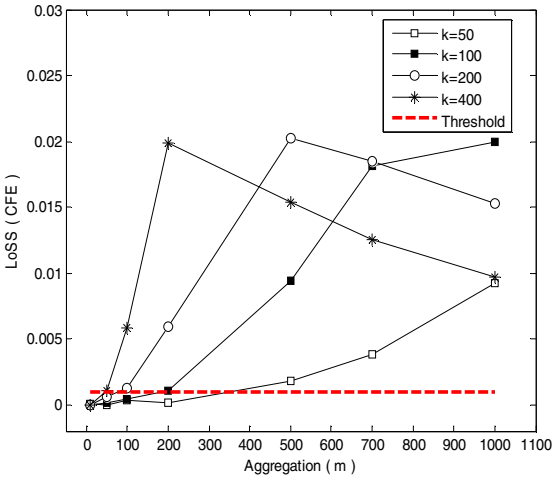


Figure 4. LoSS detection for malicious traffic FSKSMNet 2006

Table 3. Details CFE estimation for malicious: FSKSMNet 2006

Sampling (m)	Auto-correlation lag (k)					
	50	100	200	300	400	500
10	0.0001	0.0000	0.0000	0.0000	0.0000	0.0001
50	0.0000	0.0002	0.0006	0.0006	0.0011	0.0022
100	0.0003	0.0004	0.0013	0.0028	0.0058	0.0096
200	0.0002	0.0011	0.0059	0.0143	0.0199	0.0205
500	0.0018	0.0094	0.0202	0.0179	0.0154	0.0134
700	0.0038	0.0181	0.0185	0.0150	0.0125	0.0107
1000	0.0092	0.0200	0.0153	0.0118	0.0097	0.0084

Proceedings of the 4th International Conference on Information Assurance and Security (IAS 2008), pp. 192-197, Naples, Italy, September 2008.

The simulation results have shown that LoSS detection accuracy performance can be improved by incrementing the value of *m* and *k*. The drawback of using higher level of *m* is the window size packets used must be large enough in order to fulfill minimum window requirement [6] before correct estimation can be done. Meaning, more capturing time is needed when using larger sampling level such as *m* more than one second compared to micro sampling scale. Meanwhile, a bigger size of correlation lag *k* can increase CFE estimation accuracy at each value sampling value *m*. However, longer time is needed to execute larger correlation sliding process hence decrease overall estimation time process.

4. Conclusions

The results show that LoSS detection accuracy can be improved when we consider ESOSS and ASOSS models concurrently in the estimation method. Two parameters known as sampling level and correlation lag that can influence LoSS detection accuracy are investigated. The experimental results demonstrate that sampling level does not affect CFE estimation for normal traffic that follows ESOSS model. The accuracy of CFE estimation, however, is very much depending on correlation lag parameter. On the other hand, sampling level and correlation lag are identified as a prime factor that can influence LoSS detection accuracy when considering ASOSS model. The results demonstrate that LoSS can be hidden if lower sampling level and smaller correlation lag are used. Consequently, the increasing of sampling level and correlation lag values can improve LoSS detection accuracy provided the window size is sufficiently used.

4. Acknowledgments

This work was funded by Universiti Teknologi Malaysia (UTM) and Ministry of Science and Innovation Malaysia. The authors are grateful to Dr. Sulaiman Mohd Noor and Mr. Firoz for their helps in preparing the simulation of FSKSMNet dataset.

5. References

[1] W.H. Allen, and G.A. Marin, "The LoSS technique for detecting new Denial of Service attacks," *Proceedings of IEEE SoutheastCon 2004*, 26-29 March 2004, pp. 302-309.

[2] C. Cairano-Gilfedder, and R.G. Clegg, "A decade of Internet research -- advances in models and practices," *BT Technology Journal 23*, Vol. 4, Oct. 2005, pp. 115-128.

[3] M.E. Crovella, and A. Bestavros, "Self-similarity in World Wide Web traffic: Evidence and possible causes networking," *IEEE/ACM Transactions on Networking*, Vol. 5 (6), Dec. 1997, pp. 835-846.

[4] A. Erramilli, O. Narayan, and W. Willinger, "Experimental queuing analysis with long-range dependent packet traffic," *IEEE/ACM Transactions Networking*, Vol.4, 1996, pp. 209-223.

[5] A. Feldmann, A.C. Gilbert, W. Willinger, and T.G. Kurtz, "The changing nature of network traffic: scaling phenomena," *ACM Computer Communication*, Vol. 28(2), April 1998, pp. 5-29.

[6] M.Y. Idris, A.H. Abdullah and M.A. Maarof, "Iterative window size estimation on self-similarity measurement for network traffic anomaly detection," *International Journal of Computing and Information Science (IJCIS)*, Vol. 2(2), 2004, pp. 83-91.

[7] H. Kettani, "A Novel Approach to the Estimation of the Long-Range Dependence Parameter," University of Wisconsin – Madison , PhD. Thesis, 2002.

[8] S. Ledesma, and D. Liu, "Fractional Gaussian noise power spectrum synthesis using linear approximation for generating self-similar network traffic," *ACM Computer Communication Review*, Vol. 30(2), April 2000, pp. 4-17.

[9] W. Leland, M. Taqqu, W. Willinger and D. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Transactions on Networking*, Vol. 2(1), 1994, pp. 1-15.

[10] G. Mazzini, R. Rovatti, and G. Setti, "On the Aggregation of Self-Similar Processes," *IEICE Transactions Fundamentals*, Vol. E88–A (10), Oct. 2005, pp. 2656-2663.

[11] C. Park, F. Hernández-Campos, J.S. Marron, and F.D. Smith, "Long-range dependence in a changing internet traffic mix," *Computer Networks*, Vol. 48(3), Jun 2005, pp. 401-422.

[12] V. Paxson, and S. Floyd, "Wide-area traffic: The failure of Poisson modeling," *IEEE-ACM Transactions on Networking*, Vol. 3(3), June 1995, pp. 226-244.

[13] A. Qayyum, M.H. Islam, and M. Jamil, "Taxonomy of statistical based anomaly detection techniques for intrusion detection," *Proceedings of the IEEE Symposium on Emerging Technologies 2005*, 17-18 Sept. 2005, pp. 270-276.

[14] M.F. Rohani, M.A. Maarof, A. Selamat, and H. Kettani, "Loss of Self-Similarity Detection with Second Order Statistical Model and Multi-Level Aggregation Approach," *Proceedings of the International Conference on Robotics, Vision, Information and Signal Processing ROVISP2007*, Nov. 2007, pp. 152-156.

[15] W. Schleifer, and M. Mannle, "Online error detection through observation of traffic self-similarity," *IEE Proceedings on Communications*, Vol. 148(1), Feb. 2001, pp. 38-42.

Proceedings of the 4th International Conference on Information Assurance and Security (IAS 2008), pp. 192-197, Naples, Italy, September 2008.