5

HORIZONTAL LOSS ANALYSIS IN MULTI-LEVEL SAMPLING STRUCTURE

Mohd Fo'ad Rohani
Mohd Aizaini Maarof
Ali Selamat
Houssain Kettani

1. INTRODUCTION

The self-similarity model has been widely accepted to network traffic modeling and performance analysis in Local Area Network (LAN) since the scientific evidence was presented in [13], [14]. The research community's focus in network traffic modeling was shifted from assuming Poisson model which is memoryless and has smooth behavior; to assuming Long Range Dependence (LRD) and bursty behavior. Several causes have been pointed out that contribute to self-similarity behavior such as mixed behavior of TCP services model [21], mixed actions from individual users, hardware and software and heavy-tailed distribution of file transferred [4] in the network. It is important to monitor uncontrolled self-similarity structure [5], [20] and increasing LRD effect [17] in order to prevent worst Quality of Service (QoS) condition. Uncontrolled self-similarity structure will create heavy congestion at queue buffer hence drastically increase queuing delay and packet loss rate [5], [20]. On the other hand, uncontrolled increasing LRD effect will break more invariant

power law of LRD function [17] hence increasing second order QoS metric performance measurement. Therefore, Internet service providers are now faced with challenging task to ensure quality of service (QoS) of their network is uninterrupted.

Anomalous network traffic can be generated from malfunctioning network devices, network overload, Denial of Service (DoS) attacks and network intrusions [25]. Studies have shown that protocol intensity distribution plays important role to the interactions that produce self-similarity behavior [21]. Therefore, an extreme bit rate transfer from malicious packets such as DoS attacks can dominate network traffic protocol interaction intensity and produce self-similarity distribution error [2], [24]. As a result, self-similarity property is disturbed and Loss of Self-Similarity (LoSS) behavior is detected [2], [8], [15], [22], [24]. This can be used as a flag to alert security analysts of the possible presence of malicious actions. A recent technique for detecting possible presence of new DoS attacks based on LoSS without background template was presented in [2]. The method defines LoSS as self-similarity parameter or Hurst parameter (*H*) beyond normal LRD behavior using Periodogram and Whittle methods. The results have demonstrated high detection rate with an average of 60% to 84%. However, this can be assumed high false alarm detection rate and need to be improved in order to achieve a reliable and efficient online network monitoring system.

A new estimation method of *H* parameter which is more accurate and faster was developed in [10], [11]. The method was known as the Optimization Method (OM) and used Second Order Self-Similarity (SOSS) statistical model. Anomaly detection based on LoSS using SOSS and OM was suggested in [8]. Their studies, however, are limited to Hurst estimation at fixed sampling time scale. Evidences have shown that different structure of self-similarity dependence behavior can exist at different time scales as described in [1], [3], [7], [27]. Their results show that loose dependence structure (or complex scaling) exist at smaller time scale while strong dependence structure (or mono-fractal) exist at higher time scale with turning point usually associated at round trip

time (RTT). Previous works have shown that normal Internet traffic has strong dependence structure [1], [13], [14] while malicious traffic such as DoS attack tend to reduce the strength of dependence structure [17]. Therefore, this is a good indicator to propose a Multi-Level Sampling (MLS) strategy to uncover the presence of self-similarity distribution error effectively due to inappropriate sampling time scale [6], [22].

This paper presents LoSS detection analysis using horizontal approach for investigating Internet traffic behavior in the MLS structure. The method analyzes LoSS detection at each sampling level individually from lower to upper level. This can validate the normal and abnormal traffic behavior according to the exactly and asymptotically SOSS models properties. The proposed method defines LoSS detection using SOSS model and Optimization Method. The remainder of the paper is organized as follows: Section 2 presents mathematical definitions and properties of SOSS and how to estimate its parameter. Section 3 on the other hand, discusses the concept of LoSS detection and the propose Horizontal LoSS Analysis (HLA) detection. Section 4 presents our experimental procedures and the results. Finally our conclusions and future works direction are summarized in Section 5.

2. SOSS MODEL AND ESTIMATION METHOD

Let us define a second-order stationary process $X = \{X(t), t = 0,1,2,....,N\}$ with constant mean μ, finite variance σ^2 and autocorrelation function $\rho(k), k = 0,1,2,...,N$. Their definitions are given as follows:

$$\begin{aligned} &\mu = E[X(t)], \quad \sigma^2 = E[(X(t)-\mu)]^2 \\ &\rho(k) = E[(X(t)-\mu)(X(t+k)-\mu)]/\sigma^2 \end{aligned} \tag{1}$$

Let $X^{(m)} = \{X^{(m)}(t), t > 0\}$ denotes the aggregate process of X at $m = 1, 2, 3, ..., N$. Thus, for each m, $X^{(m)}$ is given by:

$$X^{(m)}(t) = \frac{1}{m} \sum_{w=m(t-1)+1}^{mt} X(w) \text{ , } t > 0 \text{ .} \tag{2}$$

Let $\gamma^{(m)}(k)$ and $\rho^{(m)}(k)$ denote the variance and autocorrelation function of $X^{(m)}$ respectively.

X is called Exactly Second Order Self-Similar (ESOSS) if $\rho(k) = \rho^{(m)}(k)$ for all $m \geq 1$. Thus, ESOSS property of correlation structure is preserved under time aggregation such that:

$$\rho(k) = \frac{1}{2}[(k+1)^{2-\beta} - 2k^{2-\beta} + (k-1)^{2-\beta}] \tag{3}$$

where $k > 0$ and $0 < \beta < 1$.

X is called Asymptotical Second Order Self-Similar (ASOSS) if $\lim_{m \to \infty} \rho^{(m)}(k) \sim \rho(k)$, $k \geq 1$ Thus, ASOSS property of correlation structure is captured such that:

$$\lim_{m \to \infty} \rho^{m}(k) \approx \frac{1}{2}[(k+1)^{2-\beta} - 2k^{2-\beta} + (k-1)^{2-\beta}] \tag{4}$$

where $k > 0$, $m > 0$ and $0 < \beta < 1$.

X is called Long-Range Dependent (LRD) if its autocorrelation function satisfies:

$$\rho(k) = ck^{-\beta} \tag{5}$$

where $k \to \infty$, $c > 0$, $0 < \beta < 1$ and the self-similarity parameter H is defined as:

$$H = 1 - \frac{\beta}{2} \tag{6}$$

There are several methods to estimate *H*. In this paper we used the Optimization Method (OM) which was developed in [10], [11] and was shown to be comparatively fast and accurate with respect to other methods in the literature such as wavelet. The method is based on how close sample autocorrelation measure fits to ESOSS model. The OM defines Curve Fitting Error (*CFE*) function $E_K(\beta)$ as in equation (7):

$$CFE = E_K(\beta) = \frac{1}{4K}\sum_{k=1}^{K}(\rho(k) - \rho_n(k))^2 \quad (7)$$

where $\rho(k)$ denotes the autocorrelation function of the model with parameter β that OM would like to fit the data to, $\rho_n(k)$ is the sample autocorrelation function of the data, k is autocorrelation lag and K is the largest of lag 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 threshold value 10^{-3}, then the data fits the model and the minimum $\hat{\beta}$ is picked to be the estimator of the parameter β [10], [11].

3. LoSS DETECTION ANALYSIS USING SOSS MODEL IN MLS STRUCTURE

In this section, we describe the fundamental of LoSS detection technique based on SOSS model and Optimization Method. Then, we propose Horizontal LoSS Analysis (HLA) method for analyzing LoSS detection in multi-level sampling structure.

3.1 LoSS Detection with SOSS Model and OM

LoSS is used as a flag to detect the possible presence of malicious actions [2], [24], [26]. The behavior of Internet traffic is considered as normal when the traffic follows the self-similarity model, otherwise it is considered as abnormal [8]. The abnormal

traffic behavior such as in the presence of DoS packets, could introduce distribution error and shifts the stationary property toward non-stationary as shown in [2], [8], [16], [17]. Data insufficient probability and detection loss probability are two important attributes that influence the correctness of anomaly detection [8]. The data insufficient probability is to identify the minimum required window size to obtain reliable self-similarity measurement, while detection loss probability is probability of detecting non-stationary data.

LoSS is detected if it fulfils two conditions which are the data must be longer than minimum window size and it must be non-stationary [8]. The experimental results in [13], [14], [18] demonstrate that windows sizes from 15-30 minutes are practical and sufficient for modern LANs Ethernet Internet traffic to comply with data insufficient probability. The self-similarity tests become more sensitive as the window size gets smaller and consequently generate false alarms if it gets too small. Therefore, in our experiments we use traces of 30 minutes in length which is above the minimum required window to fulfill data insufficient probability [8].

Based on our assumptions, we define normal behavior of self-similarity traffic as the estimated Hurst parameter H using OM is in the LRD range with $0.5 < (H, OM) < 1$ and fitting error $E_K \leq 10^{-3}$. Otherwise if $E_K > 10^{-3}$, LoSS is detected and consequently the corresponding Internet traffic is considered as abnormal. As ESOSS process has property of $\rho(k) = \rho^{(m)}(k)$ for all $m \geq 1$, it follows that ESOSS will capture the property of correlation structure that is preserved under time aggregation. Therefore, it is required to study the effect of different aggregation levels or MLS approach in order to uncover and detect the hidden self-similarity distribution error effectively. In our work, we consider sampling level at shorter time-scales [3], [6] such as $10\text{ms} \leq m \leq 1000\text{ms}$ that represent engineering factors which have stronger impact than human behaviors [3] to Internet traffic

behavior. We choose at random *m* equal to 10ms, 50ms, 100ms, 200ms, 500ms, 700ms and 1000ms.

3.2 Horizontal LoSS Analysis in MLS Structure

We proposed an approach to analyze LoSS in MLS structure known as Horizontal LoSS Analysis (HLA). The structure of Horizontal LoSS Analysis (HLA) is shown in Figure 1 and recognized as Uni-level LoSS (UNL) analysis. The HLA test is a method of level by level LoSS analysis from smaller sampling level *m* to higher sampling level *m*. The test is also known as a bottom up LoSS analysis approach for detecting traces that deviate from SOSS model. Our assumption is that the probability of normal traffic follows SOSS model at smaller *m* is much higher than larger *m* such as at $m \leq 500\text{ms}$ compared with $m > 500\text{ms}$. This is a possible reason why Hurst estimation in the previous works [8], [10], [13], [14], [18] is conducted at $m = 10\text{ms}$ or $m = 100\text{ms}$ besides the assumption of ESOSS model.

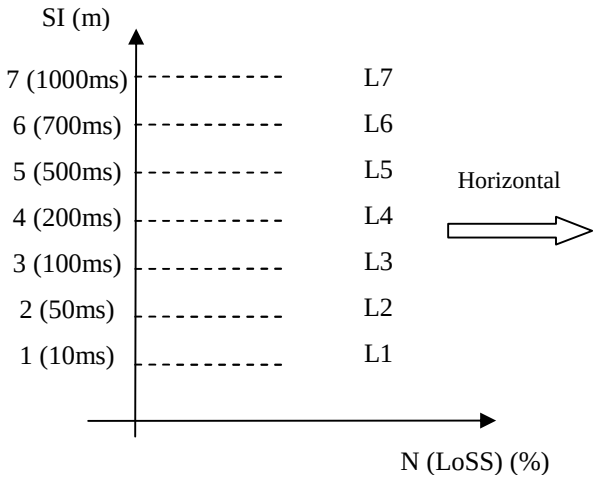

Figure 1 Horizontal LoSS Analysis (HLA)

In our analysis, we start HLA with L1 (m=10ms), and then increase to next higher level L2 (m=50ms). Similarly, the sampling level is increased to L3 (m=100ms), L4 (m=200ms), L5 (m=500ms), L6 (m=700ms) and L7 (m=1000ms) for HLA test. Figure 1 shows the structure in details. We define HLA index as percentage measurement of traffic traces that LoSS is detected for each of sampling level m separately by using equation (8).

$$HLA(\%) = \frac{N_{L(m)}}{N_T} \text{ x } 100 \tag{8}$$

where $N_{L(m)}$ represents the number of trace that LoSS is detected at sampling interval m and N_T is the total trace or slot in the detection.

The HLA index can be used as an indicator to predict Internet traffic behavior. The synthesis of HLA index based on LoSS and ESOSS leakage detection can be used to relate with normal and abnormal traffics. The possibility of Internet trace with smaller sampling level such as $m < 500$ms to exhibit ESOSS model is higher compared to larger sampling level such as $m \geq 500$ms. Thus, the increment of HLA percentage index from lower to higher m will indicate the increment of LoSS detection in the traces. This is a good sign of anomaly behavior presence in the traffic due to suspicious Internet activities or malicious packets such as DoS attacks.

There is no guideline to choose the lower and upper bound sampling level m for testing ESOSS model validation. Theoretically, there is no limit for incrementing sampling level m in the MLS structure. However, in practice we have to fulfill the minimum window requirement to estimate H correctly [8]. The constraint factors are time delay and limited buffer size especially for real time monitoring. If m is too small such as $m \leq 10$ms, then it is very difficult to reveal the self-similarity distribution error effectively [22], [23]. On the other hand, if m is too large such as $m > 1000$ms than we need longer time delay and larger buffer size before H can be estimated correctly. Therefore, the HLA test should have lower and upper bound of the incrementing sampling

level *m* so that LoSS can be analyzed accurately and efficiently. In our case, we have set lower bound of sampling level $m = 10\text{ms}$ while upper bound of sampling level $m = 1000\text{ms}$. The performance of HLA index can be used as indicator to analyze traffic behavior based on LoSS detection and ESOSS leakage criterion.

4. EXPERIMENT AND EMPIRICAL ANALYSES

The experimental works consist of two parts. The first part is to investigate the behavior of self-similarity parameters *H* and *CFE* for normal and abnormal traffic traces in different sampling levels. The second part is to analyze LoSS detection in MLS structure using HLA index for Internet traces.

4.1 Experiment Datasets

The experiments use three simulation datasets. The first is synthetic data that is generated at random by using fractional Gaussian noise (fGn) model developed in [12]. The traces theoretically inherit ESOSS property and we only consider *H* which is LRD for $0.5 < H < 1$. The length of the trace is equivalent to 15-30 minute at normal traffic capturing time for Ethernet LAN [13], [14], [18]. The second dataset is UNC2003 traces [18], [19] that are considered as normal Internet traffic activities. We use three consecutive days of the traces from Monday to Wednesday for our benchmark analysis. The third dataset is FSKSMNet Internet traffic simulation [23] on September 29, 2006 at Faculty of Computer Science and Information Systems (FCSIS) local area networks (LANs). Internet Monitoring Laboratory (InMonLab) has been setup with baseline 100BaseFX Fast Ethernet as FCSIS LANs backbone and connected to the main university Gigabit backbone. The network design is constructed with ten proxies of Virtual LANs (VLANs) for students, administrators and academic

staffs. The FSKSMnet Internet traffic simulation activities are divided into normal and abnormal traffic traces. Normal Internet activities are defined as legal Internet activities that abide with faculty network policy. On the other hand, abnormal traffic contains simulated injection of DoS flooding packets at certain rate that includes TCP Reset, TCP SYN, UDP, ICMP and IGMP flooding packets. Each capturing session is about 30 minutes. In the simulation, we simulate such that 50% of the traffic traces are contained with controlled simulation injection malicious packets using DoS generator. The normal traces are labeled as N while abnormal traces as AB.

4.2 Estimated *H* and *CFE* Analysis in MLS Structure

Figure 2 illustrates the simulation results of *H* and *CFE* parameter estimation for the fGn, UNC2003 and FSKSMnet traces at different levels of *m*. We use $CFE < 0.001$as a threshold value for the traces that fit with SOSS model. Figure 2(a), (c) and (e) demonstrate that the simulated traffic traces exhibit *H* value in the LRD domain with $0.5 < H < 1$. In normal condition, the traces follow SOSS model with estimated CFE less than a threshold. Otherwise, they deviate from SOSS model if *CFE* exceeds the threshold value as shown in Figure 2(b), (d) and (f). In our work, we use ESOSS property as LoSS detection criterion by examining the value of estimated *CFE* in the function of different sampling level *m*.

Table 1 and 2 show the average value of min, max, mean and variance of estimated *H* and *CFE* for traces $n > 1$, after considering MLS structure for each of the traces. In our experiments, we generate fGn traces with estimated average $H = 0.7$ and $\sigma^2 = 14.15\text{x}10^{-4}$. The average of estimated min and max *H* are 0.67 and 0.72. On the other hand, the normal UNC2003 traces are estimated with average estimated $H = 0.84$ and $\sigma^2 = 3.30\text{x}10^{-4}$. Their average min and max *H* are 0.81 and 0.86. This represents self-similarity parameter of modern Internet traffic

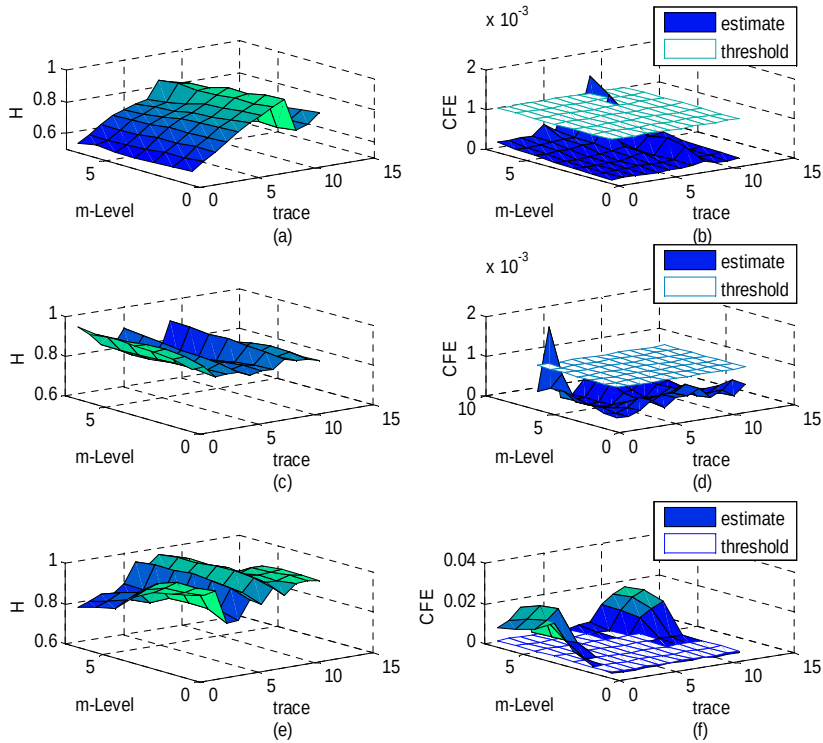

Figure 2 Hurst (*H*) estimation and Curve Fitting Error (*CFE*) at multi-level sampling (MLS) structure for data simulation: fGn- (a), (b); UNC2003 (c), (d) and FSKSMNet (e), (f)

as reported in [18]. In FSKSMNet traffic simulation, normal traffic traces show the average value of estimated $H = 0.92$ and $\sigma^2 = 3.19\text{x}10^{-4}$. Our result demonstrates that recent Internet applications have more bursty than UNC2003 traffic as shown by higher average estimated *H* value. This is illustrated by the average min and max *H* value for UNC2003 traces are 0.81 and 0.86, but for normal FSKSMNet traces are 0.89 and 0.93. The details are shown in Table 1.

Table 1 Estimated Hurst statistics (average value) in MLS structure

Trace	Average estimated *H(m)*			
	Min	Max	Mean	Variance
fGn (S)	0.67	0.72	0.70	4.15 $x10^{-4}$
UNC2003 (N)	0.81	0.86	0.84	3.30 $x10^{-4}$
FSKSMNet (N)	0.89	0.93	0.92	3.19 $x10^{-4}$
FSKSMNet (AB)	0.82	0.98	0.91	4.48 $x10^{-3}$

Table 2 Estimated CFE statistics (average value) in MLS structure

Trace	Average estimated *CFE(m)*			
	Min	Max	Mean	Variance
fGn (S)	1.25 $x10^{-5}$	3.78 $x10^{-4}$	1.50 $x10^{-4}$	3.63 $x10^{-8}$
UNC2003 (N)	1.53 $x10^{-4}$	6.36 $x10^{-4}$	3.22 $x10^{-4}$	5.73 $x10^{-8}$
FSKSMNet (N)	1.13 $x10^{-4}$	1.67 $x10^{-3}$	6.55 $x10^{-4}$	4.84 $x10^{-7}$
FSKSMNet (AB)	2.57 $x10^{-4}$	1.60 $x10^{-2}$	6.96 $x10^{-3}$	5.09 $x10^{-5}$

It is clearly shown in Table 2 that for synthetic (S) and normal (N) traces, their estimated *H* variances in MLS structure are small about 3.0 $x10^{-4}$ to 4.2 $x10^{-4}$. Similarly, their average *CFE* is always below threshold value. Their average estimated min and max *CFE* also do not exceed the threshold except for the average max *CFE* of FSKSMNet traces which is slightly higher than threshold value (i.e. $1.67x10^{-3}$). Meanwhile, for the abnormal FSKSMNet traces, the average mean *CFE* value exceeds the threshold hence demonstrating LoSS is detected and violate SOSS model. They also have larger average variances than normal traces which is approximately ten times bigger. Thus, we cannot estimate

H value correctly since the traces are not fitted with SOSS model. However, we assume that the malicious traces are trying to fit with the nearest autocorrelation curve ($\rho(k)$) and use to approximate with the nearest estimated *H* value.

With this assumption, the average *H* value for abnormal traces can be approximated with SOSS model and result in Table 2 shows the value is lower than the normal traces as reported in [15], [17]. Our result in Table 2 also illustrates that LoSS detection is uncertainty detected especially for malicious traffic. This is shown by the average min $CFE = 2.57\text{x}10^{-4}$ and average max $CFE = 1.6\text{x}10^{-2}$. The former value follows SOSS model while the later value deviates from SOSS model. Therefore, an effort to further analyze LoSS detection in multi-level structure is proposed.

4.3 LoSS Detection Analysis in MLS Structure

We simplified the *CFE* distribution for different sampling levels as graphically shown in Figure 2(b), (d) and (f) to LoSS mapping table. Table 3, 4 and 5 illustrate LoSS mapping table at different sampling level $10\text{ms} \le m \le 1000\text{ms}$ for fGn, UNC2003 and FSKSMNet simulation datasets respectively. LoSS (L) is detected if the *CFE* value is less than the threshold value otherwise SOSS (S) is detected. Then, we analyzed LoSS penetration test using HLA to investigate the validity of ESOSS property based on *CFE* criterion. The HLA detection index will be used in order to analyze LoSS detection performance to observe ESOSS leakage pattern at MLS structure.

4.4 Horizontal LoSS Analysis Approach

The purpose of HLA test is to investigate how ESOSS property is preserved in the MLS structure. We validate ESOSS property by examining traffic behavior whether it follows or violates SOSS model at each sampling level separately. We assume that for

Table 3 LoSS mapping table for fGn

SI(*m*)	Tr1	Tr2	Tr3	Tr4	Tr5	Tr6	Tr7	Tr8	Tr9	Tr10	Tr11	Tr12
10	S	S	S	S	S	S	S	S	S	S	S	S
50	S	S	S	S	S	S	S	S	S	S	S	S
100	S	S	S	S	S	S	S	S	S	S	S	S
200	S	S	S	S	S	S	S	S	S	S	S	S
500	S	S	S	S	S	S	S	S	S	S	S	S
700	S	S	S	S	S	S	S	S	L	S	S	S
1000	S	S	S	S	S	S	S	S	L	S	S	S

Table 4 LoSS mapping table for UNC2003

SI(*m*)	Tr1	Tr2	Tr3	Tr4	Tr5	Tr6	Tr7	Tr8	Tr9	Tr10	Tr11	Tr12
10	S	S	S	S	S	S	S	S	S	S	S	S
50	S	S	S	S	S	S	S	S	S	S	S	S
100	S	S	S	S	S	S	S	S	S	S	S	S
200	S	S	S	S	S	S	S	S	S	S	S	S
500	S	S	S	S	S	S	S	S	S	S	S	S
700	S	S	S	S	S	S	S	S	S	S	S	S
1000	S	L	S	S	S	S	S	S	S	S	S	S

Table 5 LoSS mapping table for FSKSMNet

SI(*m*)	Abnormal			Normal						Abnormal		
	Tr1	Tr2	Tr3	Tr4	Tr5	Tr6	Tr7	Tr8	Tr9	Tr10	Tr11	Tr12
10	S	S	S	S	S	S	S	S	S	S	S	S
50	L	S	S	S	S	S	S	S	S	S	S	S
100	L	L	L	S	S	S	S	S	S	S	S	S
200	L	L	L	S	S	S	S	S	S	L	L	L
500	L	L	L	L	S	S	S	S	S	L	L	L
700	L	L	L	L	S	S	S	L	L	L	L	L
1000	L	L	L	L	S	L	S	L	L	L	L	L

L: LoSS, S: SOSS, SI(*m*): Sampling Interval, Tr: Trace

normal traffic fGn and UNC2003 traces, they follow ESOSS model while for abnormal traffic FSKSMNet traces, they deviate from ESOSS model. We use LoSS mapping table in Table 3, 4 and 5 to perform HLA test and the result is shown in Table 6.

Table 6 HLA detection (%) for fGn, UNC2003 and FSKSMNet

Level	SI (m)	fGn (HLA%)	UNC2003 (HLA%)	FSKSMNet (HLA%)
L1	10	0	0	0
L2	50	0	0	8.33
L3	100	0	0	25.0
L4	200	0	0	50.0
L5	500	0	0	58.33
L6	700	8.33	0	75.00
L7	1000	8.33	8.33	83.33

The result in Table 6 illustrates that almost all synthetic fGn traces follow ESOSS model for $m < 700\text{ms}$. Less than 10% of the generated fGn traces demonstrate LoSS at $m \geq 700\text{ms}$. The result for UNC2003 traces shows more than 90% of the traces follow ESOSS model at $m \leq 700\text{ms}$ and less than 10% deviates from the model. This is a good indication that the probability of fGn generator to generate ESOSS model is very high at smaller sampling level such as $m \leq 500\text{ms}$. However, there is a potential for fGn to loose from self-similarity structure when m is increasing from 700ms to 1000ms, but the chance is very slim. Similarly for UNC2003 normal traces, almost more than 90% follows ESOSS model at all sampling levels m except less than 10% of the traces violate ESOSS model at higher sampling level $m = 1000\text{ms}$. The results in Table 6 show that for FSKSMNet zero LoSS occurrence is detected at $m = 10\text{ms}$. The HLA index at $m < 50\text{ms}$ is slightly increased which is less than 10%. However, the LoSS index is drastically increased from 50% at $m = 200\text{ms}$ to more than 80% at

$m = 1000\text{ms}$. The increment pattern of HLA index gives a clear warning that the FSKSMNet traces are containing a significant number of hazard traffics that effectively violate ESOSS property.

The increment of continuous HLA index pattern can be used to alert the possibility of malicious traffic presence in the unknown traffic analysis. In analyzing the HLA index pattern further, we introduce HLA sampling threshold. We assume the sampling threshold occurs at $m = 500\text{ms}$ that is middle range of the defined sampling level $10\text{ms} \le m \le 1000\text{ms}$. Figure 3 illustrates the HLA index pattern for normal and abnormal traces.

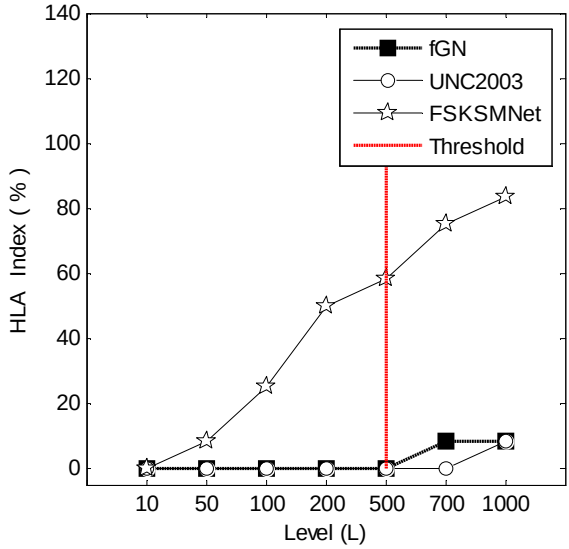

Figure 3 HLA Detection (%) for traces fGn, UNC2003 and FSKSMNet

It is clearly shown in Figure 3 that the normal traces fGn and UNC2003 have tendency to follow ESOSS model for *m* less than the sampling threshold. Meanwhile, LoSS is detected in a small percentage at the above sampling threshold that is less than 10%

for normal traces. On the other hand, the HLA index is continuously increased from 0% to more than 80% at both below and above sampling threshold for FSKSMNet traces. This indicates that FSKSMNet traces contain hazard traffic that significantly contribute LoSS and violate ESOSS property.

Our simulation result in Figure 3 also demonstrates that at typical sampling rate such as $m = 10$ms or $m = 100$ms the distribution self-similarity error has potentially to be unseen for certain malicious traffic packets. This illustrates that fixed sampling approach is not suitable to be implemented as an efficient anomaly detection method that based on LoSS. However, the performance of LoSS detection can be improved by incrementing the value of used sampling level such as above 500ms. The inconsistent occurrence of LoSS behavior at different levels of m creates a challenge to choose an optimum level that LoSS is accurately detected, if implements fixed sampling technique. Therefore as suggested in [22], the accuracy of LoSS detection can be improved if we consider a multi-level sampling approach.

5. CONCLUSION AND FUTURE WORKS

This paper investigates Internet traffic behavior based on LoSS detection using SOSS model and OM. Previous works have shown that fixed sampling is insufficient to detect LoSS accurately. Thus, a Multi-Level Sampling (MLS) approach is proposed in order to improve LoSS detection accuracy. Horizontal LoSS Analysis (HLA) is introduced to analyze LoSS at different sampling levels in the MLS structure individually. Our results show that the self-similarity leakage pattern based on LoSS detection and HLA index measurement can be used to identify normal and abnormal traffic behavior. The HLA index verifies that for the normal traffic such as synthetic fGn and UNC2003 traces are always follows ESOSS model especially at sampling level below 700ms. However, the HLA index shows LoSS is detected for both normal traces about 8% to 10% at higher sampling rate such as above 700ms. For

FSKSMNet traces that contained malicious traffic, the results show that based on HLA index, it is difficult to detect LoSS at sampling level 10ms. This is due to the hazard traffic trying to imitate normal self-similarity behavior. However, LoSS detection accuracy for the abnormal traces is increased from zero to 25% at sampling level 100ms. The results also demonstrate that the HLA index continues to increase when the sampling level is incremented. It is shown more than 50% of the FSKSMNet traces are detected with LoSS behavior at sampling level higher than 500ms. This indicates that all malicious traces in the simulated FSKSMNet traces are successfully detected at higher sampling level compared to smaller sampling level. The MLS approach is also capable to detect LoSS for legal Internet traffic activities in the FSKSMNet. The results demonstrate that the HLA index for legal traffic FSKSMNet activities is increased from zero to more than 20% at sampling level 10ms compared to higher than 500ms. Thus, the Horizontal LoSS Analysis (HLA) can be used to analyze LoSS detection and identifies Internet traffic behavior in the MLS structure. Future works will consider analysis of LoSS detection using vertical analysis that taking into account the simultaneous LoSS detection at different sampling level in the MLS structure.

REFERENCES

[1] Abry, P. and Veitch, D., “Wavelet analysis of long range dependent traffic,” *IEEE Transactions on Information Theory* 44(1), pp. 2–15, 1998.

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

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

[4] Crovella, M.E. and Bestavros, A., “Self-similarity in World Wide Web traffic: Evidence and possible causes,”

IEEE/ACM Transactions on Networking, Vol. 5, Issue 6, pp. 835 – 846, December 1997.

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

[6] Estevez-Tapiador, J.M., Garcia-Teodoro, P. and Diaz-Verdejo, J.E., "Anomaly detection methods in wired networks: a survey and taxonomy," *Computer Communications*, Vol. 27, Issue 16, pp. 1569-1584, 15 October 2004.

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

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

[9] Karagiannis, T., Molle, M. and Faloutsos, M., "Long-range dependence ten years of Internet traffic modeling;" *Internet Computing, IEEE* Vol. 8, Issue 5, pp. 57 – 64, Sept. - Oct. 2004.

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

[11] Kettani, H., and Gubner, J.A., "A Novel Approach to the Estimation of the Long-Range Dependence Parameter," *IEEE Transactions on Circuits and Systems II*, vol. 53(6), pp. 463-467, June 2006.

[12] Ledesma, S. and Liu, D., "Fractional Gaussian noise power spectrum synthesis using linear approximation for generating self-similar network traffic," *ACM Computer Communication Review*, Vol.30, no.2, pp. 4-17, April 2000.

[13] Leland, W., Taqqu, M., Willinger, W. and Wilson, D., "On the self-similar nature of Ethernet traffic," Proceedings *of ACM SIGCOMM*, Vol. 23(4), pp. 183–193, 1993.

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

[15] Li, M., "Change trend of averaged Hurst parameter of traffic under DDOS flood attacks," *Computers & Security*, Vol. 25(3), pp. 213-220, May 2006.

[16] Li, M., Jia, W. and Zhao, W., "Decision analysis of network-based intrusion detection systems for denial-of-service attacks," *Proceedings of IEEE International Conferences on Info-tech and Info-net (ICII 2001)*, Vol. 5, Beijing, PRC, pp. 1-6, 29 Oct. - 1 Nov. 2001.

[17] Owezarski, P., "On the impact of DoS attacks on Internet traffic characteristics and QoS," Proceedings of the *IEEE International Conference and Computer Communications and Networks (ICCCN'2005)*, San Diego, CA, USA, pp. 269 – 274, 17-19 October 2005.

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

[19] Park, C., Hernández-Campos, F., Marron, J. S., and Smith,F.D, "UNC DIRT Laboratory Internet traces," *http://www-dirt.cs.unc.edu/ts/* , 23 May 2003.

[20] Park, K., Kim G., and Crovella, M., "On the effect of traffic self-similarity on network performance," *SPIE International Conference on Performance and Control of Network Systems*, November 1997.

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

[22] Rohani, M.F, Maarof, M.A., Selamat, A. and Kettani, H. , "Uncovering Anomaly Traffic Based on Loss of Self-Similarity Behavior Using Second Order Statistical Model,"

International Journal of Computer Science and Network Security (IJCSNS), Vol.7 No.9, pp 116-122, September 2007.

[23] Rohani, M.F, Maarof, M.A., Selamat, A. and Kettani, H., "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*, pp. 152-15628-30 November 2007.

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

[25] Thottan, M. and Ji, C., "Anomaly detection in IP networks," *IEEE Transactions on Signal Processing*, Volume 51, Issue 8, pp. 2191 – 2204, Aug. 2003.

[26] Yan, W., Hou, E. and Ansari, N., "Anomaly Detection and Traffic Shaping under Self-similar Aggregated Traffic in Optical Switched Networks", *Proceedings of ICCTZ003*, pp. 378-381, 2003.

[27] Zhang, Z.L., Ribeiro, V., Moon, S. and Diot, C., "Small-Time Scaling Behaviors of Internet Backbone Traffic: An Empirical Study," *In IEEE Infocom, IEEE CS Press*, pp. 1826-1836, 2003.