MULTI-LEVEL SAMPLING APPROACH FOR CONTINUOUS LOSS OF SELF-SIMILARITY DETECTION USING ITERATIVE WINDOW AND STATISTICAL MODEL

MOHD FO’AD ROHANI¹, MOHD AIZAINI MAAROF¹, ALI SELAMAT¹ AND HOUSSAIN KETTANI²

¹Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia, 81300 Skudai, Johor, Malaysia

²Department of Electrical and Computer Engineering and Computer Science, Polytechnic University of Puerto Rico, P. O. Box 192017, San Juan, Puerto Rico 00919, USA

E-mail: {foad, aizaini, aselamat}@utm.my, hkettani@pupr.edu

ABSTRACT: This paper proposes a Multi-Level Sampling (MLS) approach for continuous Loss of Self-Similarity (LoSS) detection using iterative window. The method defines LoSS based on Second Order Self-Similarity (SOSS) statistical model. The Optimization Method (OM) is used to estimate self-similarity parameter since it is fast and more accurate in comparison with other estimation methods known in the literature. Probability of LoSS detection is introduced to measure continuous LoSS detection performance. The proposed method has been tested with real Internet traffic simulation dataset. The results demonstrate that normal traces have probability of LoSS detection below the threshold at all sampling levels. Meanwhile, false positive detection can occur where abnormal traces have probability of LoSS that imitates normal behavior at sampling levels below 100 ms. However, the LoSS probability exceeds the threshold at sampling levels larger than 100 ms. Our results show the possibility of detecting anomaly traffic behavior based on obtaining continuous LoSS detection monitoring.

KEYWORDS: Loss of Self-Similarity (LoSS), Multi-Level Sampling, Anomaly Traffic Detection, Second Order Self-Similarity, Iterative Window

1. INTRODUCTION

The advanced technology of attack tools and their availability on Internet have increased network vulnerability to misuse and performance traffic problems. Internet service providers are now faced with the challenging task to providing continuous network traffic monitoring system to ensure that security is well maintained. Thus, the need for reliable monitoring of network anomaly traffic detection in Ethernet network is crucial to ensure uninterrupted Quality of Service (QoS) performance. This can be achieved by continuously detecting Loss of Self-Similarity (LoSS) occurrences in network traffic when the packets are treated as a time series [1-4]. Malicious traffic such as Denial of Service (DoS) packets have tendency to contribute to the deviation from the self-similarity model [3, 4]. Consequently, LoSS is detected [1-4] and a high percentage of LoSS detection will
alert a signal of poor QoS performance [5, 6] due to the uncontrolled self-similarity structure. Implementation of LoSS detection with Second Order Self-Similarity (SOSS) statistical model has been introduced in order to attain high speed and accuracy detection needs [2]. Previous works in [1, 2] have used fixed sampling and fixed window to detect LoSS. However, fixed sampling is insufficient to reveal the self-similarity distribution error efficiently [7, 8]. On the other hand, dynamic window is needed for iterative LoSS detection in order to detect the present of anomaly traffic behavior continuously. In this work, we propose a continuous LoSS detection using iterative window and Multi-Level Sampling (MLS) approach. The Optimization Method (OM) [9, 10] is used to estimate self-similarity parameter or known as Hurst (H) value. We have evaluated the LoSS detection performance using probability of LoSS detection measurement. This paper is organized as follows: Section 2 discusses in brief the self-similarity model and the estimation method of the parameter H. The proposed LoSS detection method is discussed in Section 3 while the experimental and empirical analyses are presented in Section 4. Finally, our conclusion and future work are summarized in Section 5.

2. SOSS MODEL AND ESTIMATION METHOD

2.1 Second Order Self-Similarity Statistical Model and Parameter Estimation Method

Let define a second-order stationary process \( X = \{X(t), t > 0\} \) with constant mean \( \mu \), finite variance \( \sigma^2 \) 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 \( X^{(m)} = \{X^{(m)}(t), t > 0\} \) denote the aggregate process of \( X \) at aggregation level \( m > 0 \).

Thus, we have:

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

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 \( m \geq 1 \). In ESOSS, 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

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

where \( k > 0, \ m > 0 \) and \( 0 < \beta < 1 \). \( X \) is called Long-Range Dependent (LRD) if its autocorrelation function satisfies: \( \rho(k) = ck^{-\beta} \) 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 [9, 10] 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:

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

(6)

where $\rho(k)$ denotes the autocorrelation function of the model with parameter $\beta$ 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 $\beta$ [9].

### 2.2 Loss of Self-Similarity Detection Model

Let $X(t)$ be a stochastic time series data with second order stationary property. The autocovariance decay of $X(t)$ and aggregated $X^{(m)}(t)$ that follow ESOSS model can be written as in equation (7):

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

(7)

where $m$ is sampling level, $k$ is correlation lag, $C_0$ is constant and $\beta$ is self-similarity parameter. Equation (7) shows sampling level $m$ does not play an important role in the autocorrelation process for exactly self-similarity model. Nevertheless, the self-similarity processes can also be treated as processes $x(j)$ in the class $X$ of those stationary processes that feature an asymptotic decay in autocovariance [11]. Thus, we should take into account ESOSS and ASOSS models concurrently in order to estimate the self-similarity parameter for normal and abnormal traffic correctly. To develop LoSS detection using both models; let denotes autocovariance, variance and autocorrelation for aggregated process $X^{(m)}(t)$ as shown in equation (8), (9) and (10):

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

(8)

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

(9)

$$\lim_{m,k \to \infty} \rho^m(k) = \lim_{m \to \infty} \left( \frac{\gamma^m(k)}{\gamma^m(0)} \right) \sim \frac{C_1 m^{-\beta} k^{-\beta}}{C_2 m^{-\beta}} \sim C_3 k^{-\beta}$$

(10)

where $C_1$, $C_2$ and $C_3$ are constants. Previous works in [9, 12, 13] have assumed normal Internet traffic always follow ESOSS model where its characteristics can be described by equations (8) to (10). It is clearly shown in equation (10) that autocorrelation decay is not influenced by aggregation ($m$) parameter for normal traffic. In ESOSS model, the correlation lag ($k$) plays an important role to obtain high accuracy self-similarity parameter ($\beta$) estimation.
A high density of DoS packet, such as in the presence of malicious traffic DoS traffic, can produce self-similarity distribution error and disturbs Internet traffic behavior. The normal characteristics of equations (8), (9) and (10) are disrupted. Consequently, LoSS is detected. Equation (11) shows that for abnormal traffic, the autocovariance and variance decay of \( C_m \beta^k \) and \( (C_{2m} \beta^k) \) are not identically distributed. As a result, the characteristic of equation (11) deviates from normal self-similarity pattern as shown in equation (10).

\[
\lim_{m,k \to \infty} \rho^m(k) = \lim_{m,k \to \infty} \left( \frac{\gamma^m(k)}{\gamma^m(0)} \right) \sim \left( \frac{C_m \beta^k}{(C_{2m} \beta^k)} \right) k^{-\beta} \neq C_k k^{-\beta} \tag{11}
\]

Equations (10) and (11) demonstrate that aggregation \( m \) and correlation lag \( k \) are two parameters that need to be considered for estimating the CFE (or \( E_k(\beta) \)) value in equation (6) correctly in order to improve LoSS detection accuracy. Previous work in [2] had used OM to estimate self-similarity parameter to detect anomaly traffic. Their method is based on LoSS detection and their definition of normal traffic behavior as follows:

\[
H_L \cap CF_N \tag{12}
\]

where \( H_L : H \in [0.5,1] \) and \( CF_N : CFE < \text{Threshold} \) (at fixed sampling \( m \)).

On the other hand, abnormal behavior defined as LoSS is detected at normal fixed sampling such as \( CF_N \) where \( CFE > \text{Threshold} \). Nevertheless their technique has only considered fixed sampling such as 10ms or 100ms which is not sufficient to reveal the hidden self-similarity distribution error accurately. Uncovering LoSS at multi-level sampling (MLS) time scale was suggested in [7, 8] in order to reveal the hidden self-similarity distribution error effectively. We investigate multi-level sampling at micro sampling range that is below one second which known as engineering factor [14] for Internet protocol design. We limit the value of correlation lag below five hundred in order to avoid longer time estimation in order to maintain high speed LoSS detection performance.

3. CONTINUOUS LOSS DETECTION USING ITERATIVE WINDOW AND MULTI-LEVEL SAMPLING APPROACH

We have considered a multi-level sampling approach with sampling level \( m \) in the range of \( 10ms \leq m \leq 1000ms \) in order to investigate the proposed LoSS detection method [7, 8, 14]. We define iterative window as the window size is continuously incremented in a defined fixed size window denoted by \( \Delta W \). Each of packet traces that accumulated at the current iterative window buffer is then being evaluated for LoSS detection occurrence. The proposed LoSS detection method has considered ESOSS and ASOSS models with their properties are described in equations (7), (10) and (11). Meanwhile, the normal and abnormal traffic behavior has implemented definition based on equation (12) for each of fixed sampling traces in the MLS structure. The procedure of continuous LoSS detection method consists of two processes that includes initializing
window and LoSS detection process. The initialization window is done to fulfill the minimum window size requirement before correct estimation of $H$ [2] can be obtained. This involves window size that meets the CFE criterion below the threshold value. In addition, the definition of initialization window also considers normal LRD behavior such as shown in equation (12). The process of LoSS and SOSS detection can only be continued if the initialization window has been established. However, if initialization window fails (IF) even though enough capturing time is given such as 30 minutes as used in [12, 13], we declare malicious traffic behavior is captured. The algorithm for initialization window process is shown in Fig. 1.

$$\text{set stepSize, } \Delta W$$
$$\text{set window, } W = W + \Delta W$$
$$\text{while } (W < W_{\text{max}})$$
$$\quad \text{estimate } H \text{ and CFE}$$
$$\quad \text{if } (\text{CFE} < \text{Threshold}) \& \& (0.5 < H < 1)$$
$$\quad \\
$$\quad \quad \text{Initialization Window success}$$
$$\quad \quad \text{SOSS is detected}$$
$$\quad \quad \text{proceed with LoSS detection}$$
$$\quad \quad \text{else}$$
$$\quad \quad \text{increment } \Delta W++$$
$$\quad \quad \text{set } W = W + \Delta W$$
$$\quad \text{end}$$
$$\text{end}$$
$$\text{if } (W \geq W_{\text{max}}) \& \& (\text{errCheck} = 0)$$
$$\quad \text{Initialization Window failed}$$
$$\quad \text{suspect suspicious behavior}$$
$$\text{end}$$

Fig. 1: Initialization window process.

On the other hand, continuous LoSS and SOSS detection with iterative window is based on whether $\text{CFE} > 10^{-3}$ for LoSS and $\text{CFE} \leq 10^{-3}$ for SOSS. Similar to the process of initialization window procedure, the normal self-similarity behavior that follow SOSS model is also taken into account the normal LRD behavior as defined in equation (12). The LoSS detection algorithm is shown in Fig. 2.

$$\text{while } (W < W_{\text{max}})$$
$$\quad \text{increment } \Delta W++$$
$$\quad \text{update } W = W + \Delta W$$
$$\quad \text{estimate } H \text{ and CFE}$$
$$\quad \text{if } (\text{CFE} < \text{Threshold}) \& \& (0.5 < H < 1)$$
$$\quad \quad \text{SOSS is detected}$$
$$\quad \quad \text{else}$$
$$\quad \quad \text{LoSS is detected}$$
$$\quad \text{end}$$
$$\text{end}$$

Fig. 2: LoSS and SOSS detection process.

We introduce measurement probability of LoSS detection to assess the effectiveness of the proposed detection algorithm. Thus, we define iterative window update in a continuous
hunting mode as \( W_i = \{W_{i1}, W_{i2}, W_{i3}, \ldots, W_{iN}\} \) for \( i = 1, 2, 3, \ldots, N \). For each of the updated window \( W_i \), if LoSS is detected then we update LoSS window equal to \( W_i \) or else update SOSS window with \( W_i \). Suppose we have the updated LoSS and SOSS window as follows: 

\[
\text{LoSS}(i) = \{L_1, L_2, L_3, \ldots, L_K\} \quad \text{for} \quad i = 1, 2, 3, \ldots, K \\
\text{SOSS}(j) = \{S_1, S_2, S_3, \ldots, S_M\} \quad \text{for} \quad j = 1, 2, 3, \ldots, M.
\]

Then, we define the probability of LoSS \((P_L)\) and the probability of SOSS \((P_S)\) detection using equations (12) and (13):

\[
P_L = P(\text{LoSS}) = \sum_{i=1}^{K} \frac{L_i}{L_j + S_k}, \quad (12)
\]

\[
P_S = P(\text{SOSS}) = \sum_{j=1}^{M} \frac{S_j}{L_j + S_k}, \quad (13)
\]

There are three possible conditions for \( P_L \) and \( P_S \) value when we change the sampling level value from lower \( m \) to higher \( m \) as follows:

i) \( P_L < P_S \)

ii) \( P_L = P_S \)

iii) \( P_L > P_S \)

The first condition represents the normal behavior of Internet traffic that dominated by self-similarity model and we refer this as near to normal. The smaller value of \( P_L \) indicates the higher probability of Internet traffic follows the normal SOSS model. The critical point occurs when both probability values are the same. At this point, LoSS and SOSS occurrences are equally detected and we refer this as critical point where \( P_L \) and \( P_S \) equal 0.5. On the other hand, the traffic behavior is treated as near to abnormal when the value of \( P_L \) is larger than \( P_S \) which shows self-similarity leakage is dominant. This condition is usually occurring at higher sampling level such as \( m \) is larger than 500 ms. At this condition, we can predict that the traffic is deviated from normal SOSS model and can be treated as contains suspicious or malicious packets traffic such as DoS activities.

4. EXPERIMENTS AND EMPIRICAL ANALYSES

We have simulated the FSKSMNet Internet traffic traces collected on September 29, 2006 at Faculty of Computer Science and Information Systems (FCSIS) [7] in order to evaluate the proposed LoSS detection method. The network infrastructure consists of ten VLANs with 100BaseFX Fast Ethernet backbone which is connected to university Gigabit backbone. The simulation is divided into normal and abnormal traffic. Normal traffic is defined as Internet activities that strictly follow FCSIS network policy. On the other hand, abnormal traffic contains simulated injection of DoS flooding packets at controlled rate. Each of simulation traces is about 30 minutes and the details are shown in Table 1. FNet-1 and FNet-2 are normal traces while FNet-3 contains UDP, TCP SYN and TCP RST, while FNet-4 has TCP SYN and UDP flooding packets. In the experiments, we set correlation lag \( k \) equal 100 in order to ensure the elapse time for correlation process in the OM.
remains fast and optimize. If we set higher value of \( k \), we need more sample size packets at the higher sampling level \( m \) in order to fulfill minimum window requirement as required in the proposed LoSS detection method. In the parameter estimation using OM, the threshold value for estimated \( CFE \) is set to \( 10^{-3} \) to obtain correct estimation \( H \) as used in [2, 9, 10]. We choose sampling level randomly equal 10ms, 50ms, 100ms, 200ms, 500ms, 700ms and 1000ms [7, 8, 14] to represent multi-level sampling structure in the experiments.

Table 1: FSKSMNet Dataset 2006 [7]

<table>
<thead>
<tr>
<th>Trace</th>
<th>Total Packet</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FSKSMNet-Normal</strong></td>
<td></td>
</tr>
<tr>
<td>FNet-1</td>
<td>IP=4197509: TCP(97.87%), UDP(1.69%), ICMP(0.12%), IGMP(0.01%), Others(0.31%)</td>
</tr>
<tr>
<td>FNet-2</td>
<td>IP=7371721: TCP(92.17%), UDP(0.93%), ICMP(0.07%), IGMP(0.004%), Others(6.83%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trace</th>
<th>Total Packet</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FSKSMNet-Abnormal</strong></td>
<td></td>
</tr>
<tr>
<td>FNet-3</td>
<td>IP=7468026: TCP(85.60%), UDP(14.35%), ICMP(0.04%), IGMP(0.01%), Others(0.005%)</td>
</tr>
<tr>
<td>FNet-4</td>
<td>IP=9707011: TCP(69.17%), UDP(30.77%), ICMP(0.04%), IGMP(0.005%), Others(0.02%)</td>
</tr>
</tbody>
</table>

The results of continuous LoSS detection for FSKSMNet traces are shown in Fig. 3 and Fig. 4 respectively. Fig. 3 illustrates LoSS detection for normal traces while Fig. 4 for malicious traces. Our simulation result shows zero probability of LoSS is detected at all sampling levels of \( m \) for normal trace FNet-1. The result demonstrates that the normal FNet-1 trace strictly follows ESOSS model as clearly shown in Fig. 3(a). However, there is a tendency for normal Internet activities that produce traffic behavior with a small portion of LoSS detection occurs at higher value of \( m \) such as above 500ms. This can be illustrated by normal trace FNet-2 in Fig. 3(b). Meanwhile, our result also demonstrates that LoSS is hardly detected for malicious traces FNet-3 at small value of \( m \) such as lower than 100ms as shown in Fig. 4(a). However, the LoSS occurrence is revealed clearly at higher value of \( m \) such as larger than 100 ms. The results show that multi-level sampling approach can improve LoSS detection performance as illustrates in Fig. 3 and Fig. 4. The finding is agreed with previous results in [7, 8] where LoSS detection model is implemented by using the property of ESOSS and ASOSS models such as shown in equations (7), (10) and (11).
Another indication for malicious or suspicious behavior presence in the traces is that the failure to obtain the initialized window as illustrated by abnormal trace FNet-4 in Fig. 4(b). As shown in Fig. 4(b), the algorithm is not able to obtain the initialized window or window Initialization Failed (IF) when the value of $m$ is larger than 100 ms. This is an additional alert signal that reveals the presence of abnormal traffic packets in the network despite the capability to imitate SOSS model behavior at lower sampling such as value of $m$ is less than 100 ms.
The probability of continuous LoSS detection for normal and abnormal traffic is shown in Fig. 5. The LoSS probability for the normal FNet-1 trace is less than 0.1 for all values of $m$, which is much smaller than the threshold 0.5. Similarly, the normal FNet-2 trace has probability of LoSS value less than 0.1 at sampling level $m$ below 500 ms but increase slowly to 0.4 at higher value of $m$ larger than 500 ms. However, the probability of LoSS for malicious traces FNet-3 and FNet-4 have similar patterns with FNet-2. They hide the self-similarity distribution error at value of $m$ below 100 ms but the error is exposed clearly at value of $m$ larger than 100 ms. The result in Fig. 5 also illustrates that both malicious traces FNet-3 and FNet-4 have exceeded the critical probability of LoSS threshold at 0.5. This will give a clear alarm signal to network security analyst that a severe ESOSS leakage is occurred during the continuous LoSS detection process. We
define ESOSS leakage as in any level of \( m \); there is at least one sampling level where LoSS can be detected. Another word, the self-similarity leakage demonstrates the breach indication of ESOSS model property. Our assumption is that if probability of LoSS value estimated below the threshold i.e. \( P_L < 0.5 \), then the traffic behavior will be dominated by SOSS model. Otherwise, the majority of updated windows are detected with LoSS which alerts a large amount of ESOSS leakage occurrence. This can be shown in Fig. 5 where FNet-3 and FNet-4 have critical ESOSS leakage warning compared to FNet-1 and FNet-2 at sampling level \( m \) larger than 100ms. Thus, the probability of LoSS measurement can be used as a technique to reveal the abnormal traffic behavior in a continuous LoSS detection at multi-level sampling approach as illustrates in Fig. 5.

![Fig. 5: Probability of continuous LoSS detection.](image)

5. CONCLUSION AND FUTURE WORK

In this paper, we proposed a multi-level sampling approach for continuous Loss of Self-Similarity (LoSS) detection monitoring using iterative window technique. The LoSS detection method is based on Second Order Self-Similarity Statistical (SOSS) model while the Optimization Method (OM) is used to estimate the self-similarity parameter. The proposed method uses CFE criterion and LRD behavior to detect LoSS. If the estimated CFE is not exceeding threshold value and \( H \) is in the range of LRD, then it is identified as SOSS detection. Otherwise, LoSS is detected. The probability of LoSS introduced in this work can continuously measure LoSS detection performance. Our simulation results show that at all sampling levels, normal traces is usually exhibit small probability of LoSS detection compared to probability of SOSS detection. On the other hand, there is a possibility for malicious traces to have probability of LoSS detection that imitate normal traffic behavior which normally occur at lower sampling level such as below 100ms. However, the probability of LoSS detection for abnormal traffic is clearly exposed when sampling level is larger than 100 ms. This is a promising approach to develop an improved continuous LoSS detection for anomaly traffic behavior monitoring based on SOSS model.
and multi-level sampling approach. In future, we plan to implement continuous LoSS detection method using average value of multi-level sampling parameters such as average CFE and $H$ in multi-level sampling structure. Moreover, the approach can verify whether Internet traffic traces are preserving or violating the self-similarity property based on ESOSS and ASOSS models in order to develop a reliable network traffic monitoring systems.

ACKNOWLEDGEMENT

This work was funded by Universiti Teknologi Malaysia (UTM) under the Human Resource Development Program. The authors are grateful to Dr. Sulaiman Mohd Noor at CICT, UTM and Mr. Firoz at Unit IT, FSKSM for their help in preparing the simulation of FSKSMNet 2006 datasets.

REFERENCES


