Iterative Window Size Estimation on Self -Similarity Measurement for Network Traffic Anomaly Detection
Mohd Yazid Idris, Abdul Hanan Abdullah and Mohd Aizaini Maarof
Pages 84 - 91

Iterative Window Size Estimation on Self-Similarity Measurement for Network Traffic Anomaly Detection

Mohd Yazid Idris Abdul Hanan Abdullah Mohd Aizaini Maarof
Faculty of Computer Science and Information System
Universiti Teknologi Malaysia
81310 UTM Skudai, Johor, Malaysia
{yazid,hanan,maarofma}@fsksm.utm.my

Abstract: *An iterative method for estimating the optimum sample time (or simply window size) in self-similarity measurement of network traffic is introduced. The main purpose of this measurement is to identify anomaly in network traffic. When the network traffic is close to the self-similarity model, it is considered as normal while otherwise it is not. Since, this model is related to a long-range dependence process, providing data in long period of time will increase the closeness of the network traffic towards the model. On the other hand, increasing the time range is one of the factors that will increase detection loss probability where an intrusive pattern may hide inside the normal data. Thus, the purpose of this method is to minimize the curve-fitting error on self-similarity measurement and detection loss probability in anomaly detection. This iterative method was applied to network traffic data provided by Lincoln Lab, Massachusetts Institute of Technology (MIT). The result has shown, that this method is able to estimate an optimum window size that is capable to reduce detection loss probability and maintain a low error rate.*

Keywords: *Iterative method, window size, network traffic, anomaly detection, self-similarity model.*

Received: *September 13, 2004* | **Revised:** *February 10, 2005* | **Accepted:** *March 18, 2005*

1. Introduction

In most general terms, a self-similar pattern can be defined as repeated patterns or shapes of different sizes. This pattern can be modeled in statistical processes such as fractional Gaussian noise (fGn) and fractional autoregressive integrated moving average (farima). These patterns were also discovered in several different data types including network traffic data.

Initially, research on self-similarity in network traffic was done by Leland [9]. He conducted a series of observations on high quality network traffic and found that network traffic exhibits the features of long-range dependence and self-similarity. Self-similarity in network traffic eventually attracted many other researchers to do research specifically in wide area networks and local area networks using different network protocols such as transfer control protocol (TCP), internet protocol (IP), hypertext transfer protocol (HTTP), [8, 13] and also research in network traffic anomaly detection [4, 5, 6, 7].

Anomaly detection in network traffic is one of the most challenging research areas. Previously there existed no specific distributions in order to model the network traffic data. Thus, it is difficult to distinguish statistically between a normal and an abnormal pattern in network traffic [10, 14]. The Poisson model, for example, will fail in network traffic modeling, since the distributions tend to be uniform when the number of network packets increase.

There are several studies of anomaly detection in network traffic using different self-similarity measurement methods such as *Rescale Adjusted Range (R/S), Periodogram* [4, 5, 6] and the *Wavelet* method [7]. One of the important parameters in this measurement is the time range during which the data are sampled, the so-called “window size”. The longer the window size is, the more reliable the measurement becomes [1]. Otherwise, it will become unreliable when we increase the size in order to detect intrusive pattern in network traffic. The estimation of window size was not previous research as shown in literature; researches used predetermined fix sample sizes.

A current study on self-similarity measurement used an optimization method that made this window size estimation more realistic because of the increased calculation speed. In addition, this method also provides a technique to identify whether the data tend toward the self-similarity model according to the curve-fitting error value calculated. The error percentage can be changed in order to select the optimum window size.

The rest of this paper is organized as follows: Section two is an introduction to the self-similar process and optimization method. Section three is about the iterative window size estimation method, section four is about experimental procedures and section five presents the results. Finally, section six contains our concluding remarks.

2. Preliminaries

2.1. Self-Similar Process

In order to explain the self-similar process, let X_i be a discrete stochastic process and $i, j, k \in Z$ where Z are integer values. Process X_i is self-similar process if the process holds the properties below:

- The covariance of X_i decrease slowly
- The autocorrelation of X_i decay in hyperbolic form
- It has a Hurst effect where the Hurst parameter, *H*, and 0.5 < *H* <1.0.

Auto-covariance will lead to slow decaying and formation of a hyperbolic auto-correlation when X_i are at least weak stationary. X_i is considered as weak stationary when it satisfies two conditions [1, 3]. First, its mean, μ, must satisfy equation (1):

$$\mu(X_i) = \mu(X_j) \tag{1}$$

Secondly, the auto-covariance, γ, must satisfy equation (2):

$$\gamma(X_i, X_j) = \gamma(X_{i+k}, X_{j+k}) \tag{2}$$

Auto-covariance of the process is defined as equation (3):

$$\gamma(X_i, X_j) = \mu[(X_i - \mu(X_i))(X_j - \mu(X_j))] \tag{3}$$

While, auto-correlation function, ρ, for the process is defined as equation (4):

$$\rho(X_i, X_j) = \frac{\gamma(X_i, X_j)}{\sigma^2} \tag{4}$$

Where the variance, $\sigma^2 = \gamma(0)$.

There are two types of self-similarity. It is either asymptotically or exactly self-similar. Let *m* be an aggregation level of X_k. $X_k^{(m)}$ denotes the average series of *X* over non-overlapping windows of size m given in equation (5) below:

$$X_k^{(m)} = \frac{1}{m}(X_{(k-1)m} + .. + X_{km-1}), k \geq 1 \tag{5}$$

When auto-correlation of $X_k^{(m)}$ satisfies $\rho^{(m)}(k) \to \rho(k)$, the process is identified as asymptotically self-similar. Otherwise, if $\rho^{(m)}(k) = \rho(k)$ the self-similarity is completely self-similar.

2.2. Optimization Method

There are several methods that previously described how to measure self-similarity. One of the reliable methods widely used by the networking community is the wavelet method.

However, the wavelet method is complicated and time-consuming [1]. Recently, an optimization method was introduced to increase the accuracy and to reduce measurement time. Results have shown that this method is more accurate and faster than the wavelet method.

Due to its simplicity and accuracy, the employment of this method has become more interesting in network traffic anomaly detection. Clearly, this method benefits a detection method that needs iterative measures. Thus, the employment of this method in this research is justified.

Technically, this method is based on the curve-fitting error between the auto-correlation curve of sample data $\hat{\rho}(k)$ and the autocorrelation function $\rho(k)$ model defined as in equation (6):

$$\rho(k) = \frac{1}{2}((k+1)^{2H} - 2k^{2H} + (k-1)^{2H}), k \geq 1 \tag{6}$$

The curve-fitting error is measured using error function $E_K(H)$, where *H* is Hurst parameter and $E_K(H)$ is defined as in equation (7):

$$\frac{1}{4K} \sum_{k=1}^{K} \{\rho(k) - \hat{\rho}(k)\}^2 \tag{7}$$

For the minimum point of the curve-fitting error, the following must hold: $E_K(H) \leq 10^{-3}$ [1, 3]. Then the process follows the appropriate auto-correlation model with Hurst parameter, *H* and considered as stationary data. Otherwise, if $E_K(H) > 10^{-3}$ it indicates that the provided sample data are not sufficient or it does not follow the model; thus parameter estimation is unreliable. It also indicates the provided data are non-stationary.

Figure 1 shows a sample of H where H is the minimizer of the error function. The minimum point of the curve-fitting error is smaller than 10^{-3} where $E_K(H) = 5.1698 \times 10^{-4}$ with H estimated equal to 0.78.

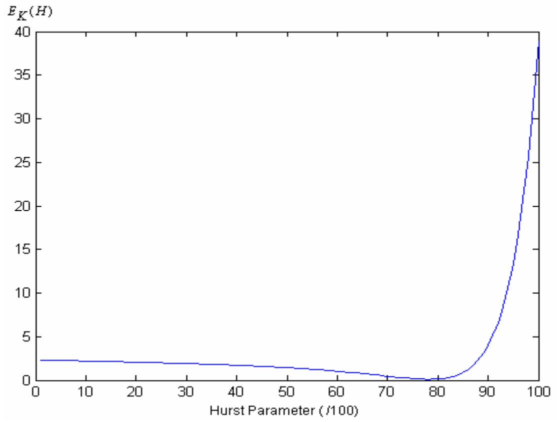


Figure 1. Curve-fitting error graph.

3. Iterative Window Size Estimation

The main objective of this estimation is to derive an optimum window size. For the first step, one needs to observe the impact of increasing window size on the probability of the data insufficiency. This method will be explained in Section 3.1.

However, increasing window size without considering the intrusive pattern duration may cause loss of detection. Thus, estimating detection loss probability becomes an important part in this window size estimation as described further in Section 3.2.

Finally, from differences of both estimated probabilities the optimum window size will be derived as in Section 3.3.

3.1. Iterative Estimation of Data Insufficient Probability

The data insufficient probability is important to identify how short the window size may be made in order to make the self-similarity measurement reliable. The data insufficient probability, P_e, is defined as a total number of non-stationary data, ε, divided by total number of samples or windows, τ, within fix window size, ω as in equation (8):

$$P_e = \frac{\varepsilon}{\tau} \tag{8}$$

Where, τ is defined as equation (9), with λ the length of data in the time unit of seconds.

$$\tau = \left\lfloor \frac{\lambda}{\omega} \right\rfloor \tag{9}$$

The estimation of this probability is done iteratively n times using n different window sizes with a fixed step size. Figure 2 illustrates four different window sizes with a step size of 100 seconds:

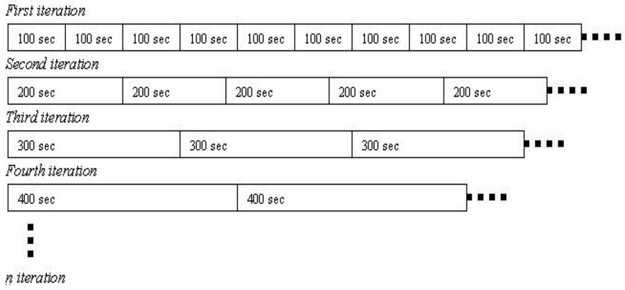


Figure 2. Window sizes and iterations.

3.2. Iterative Estimation of Detection Loss Probability

Detection loss occurs when the detectors fail to detect an intrusive pattern. In order to detect such a pattern, it first needs to consider if the intrusive network traffic data are non-stationary data. Otherwise, the detection becomes impossible and deviation between normal and abnormal pattern cannot be identified using this technique. The normal template of the Hurst parameter where $0.5 < H < 1.0$ is proven difficult since this parameter is unpredictable and have a large variance for each sample on the normal data [5].

This estimation was implemented on data merged between from non-stationary and stationary data steams. For each window the non-stationary data are extracted from their original location and merged with an attack-free background. The fix attack data were then merged with increased normal data in different window sizes. For each window size we sampled τ times and the count total number detected, υ. We defined the detection rate as in equation (10):

$$P_{\det} = \frac{\upsilon}{\tau} \tag{10}$$

Where the detection loss is defined as in equation (11):

$$P_{los} = 1 - P_{\det} \tag{11}$$

3.3. Estimation of Window Size

In order to estimate the optimum window size we consider two parameters, data insufficient probability,

P_{err} and detection loss probability, P_{los}. Our goal was to find the minimum point of the difference between those two parameters, P_{dif}. At this minimum point the data insufficient probability as well as detection loss probability are low. Thus, optimum window size, ω_{opt}, can be defined as a minimizer of P_{dif} such as in equation (12):

$$P_{dif}(\omega_{opt}) = \min_{i=1}^{n} \left\{ \left[P_{err}(i) - P_{los}(i) \right]^2 \right\} \tag{12}$$

Where, *i*: number of window sizes.

3.4. Window Size Iterative Estimation Algorithm

The algorithm is divided into three sections according to the methods described before. All sections below sketched as algorithms:

i. Estimation of data insufficient probabilities.

Input: attack free network traffic data
Output: data insufficient probabilities

```
procedure data-insufficient-probability
begin
   window-size, $\omega$ = initial-window-size;
   length, $\lambda$ = size of data;

  for window-size = initial-window-size to max-
        iteration step window-step-size
        window-number = floor(length/window-size);
        for window = 1 to max-window-number, $\tau$
              $\varepsilon$ = count ( $E_K(H) > 10^{-3}$ ) ;
        end
        $P_{err}$ (window-size) = $\varepsilon/\tau$ ;
        window-size = window-size + window-step-size;
    end
end
```

ii. Estimation of detection loss probabilities

Input: merged network traffic data between normal and fix intrusive data
Output: detection loss probabilities

```
procedure detection-loss-probability
begin
   window-size, $\omega$ = initial-window-size;
   length, $\lambda$ = size of data;

  for window-size = initial-window-size to max-
        iteration step window-step-size
        window-number = floor(length/window-size);
        for window = 1 to max-window-number
              $\upsilon$ = count ( $E_K(H) > 10^{-3}$ ) ;
        end
        $P_{det} = \upsilon/\tau$ ;
        $P_{los}$ (window-size) = $1 - P_{det}$ ;
        window-size = window-size + window-step-size;
    end
end
```

iii. Estimation of minimum probability differences

Input: error and detection loss probabilities
Output: optimum window size, ω_{opt}

```
procedure window-size-estimation
begin
   window-size, $\omega$ = initial-window-size;
   for window-size = initial-window-size to max-
         iteration step window-step-size
      $P_{dif}$ (window-size) = { $P_{err}$ (window-size) -
                              $P_{los}$ (window-size) }$^2$;
      window-size = window-size + window-step-size;
  end

  $P_{ave}$ = average ( $P_{dif}$ ) ;
 if min( $P_{ave}$ ) then $\omega_{opt}$ = window-size ;
end
```

ω_{opt} is the ultimate result in this method and algorithm, which is derived from the smallest differences between the insufficient data probability and the detection loss probability.

4. Experimental Procedures

In this section, three main procedures will be explained. First, the data preparation followed by the experimental data and finally, the implementation.

4.1. Data Preparation

Network traffic data can be dumped easily at any point of the network segment. One of the constraints on those types of traffic data is the difficulty to ensure whether the data are normal. Thus, for accuracy reasons and research purposes, synthetic network traffic data are needed. There are several synthetics network traffic data sets currently available. The

Lincoln Laboratory, Massachusetts Institute of Technology (MIT) provides one of them; it is known as MIT/DARPA data sets [2]. These data sets are commonly used for evaluating network intrusion detection.

The MIT/DARPA data sets are provided for the years 1998 to 2000. Each year contains several weeks of data sets. The data sets can be classified as training data sets and evaluation data sets. In this experiment, we restricted our analysis to the year 1999, first week data. These data sets are categorized as normal network traffic data sets from Monday to Friday. Each data set comprises of several information packet headers.

We are interested in the number of packets received and transmitted every second or identified as traffic load per second. Thus, the data should be extracted and analyzed based on the time stamp provided. A sample of plotted data is shown in Figure 3.

4.2. Experimental Data

There are two different types of data in this experiment namely normal and merged network traffic data. Normal data are extracted from first week 1999 MIT/DARPA data sets from Monday to Friday. Merged data contains fix duration intrusive data with normal background data sets. These normal background data sets are the same as normal data sets that start from Monday to Friday. The intrusive network traffic data are pre-identified and extracted from the fifth week of MIT/DARPA data with fix duration and curve-fitting error value. Furthermore, the intrusive data must be non-stationary data. For experimental purposes, eighty different data sets are provided. Five data sets are normal data sets and seventy-five data sets are merged data sets. The number of merged data sets depends on the number of window sizes that will be implemented. Thus, each window size will have its own merged data set. For example, Figure 4 shows merged data set for window size 2000 seconds.

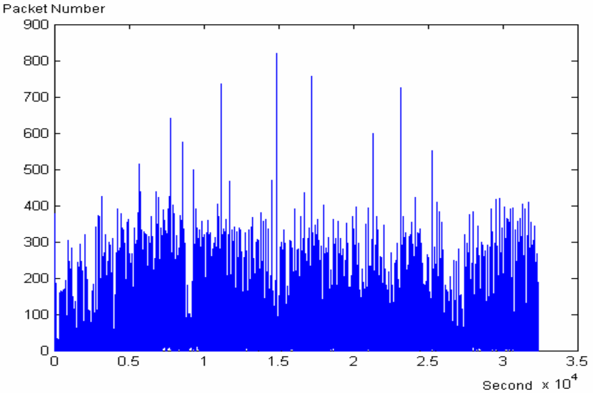


Figure 3. Packet number / second on Monday first week.

Each data set contains on average twenty-two hours of network traffic data. However in this experiment, we try to reduce the low network usage effect such as after working hours that can affected the estimated window size. Thus, the data are restricted on the first nine hours.

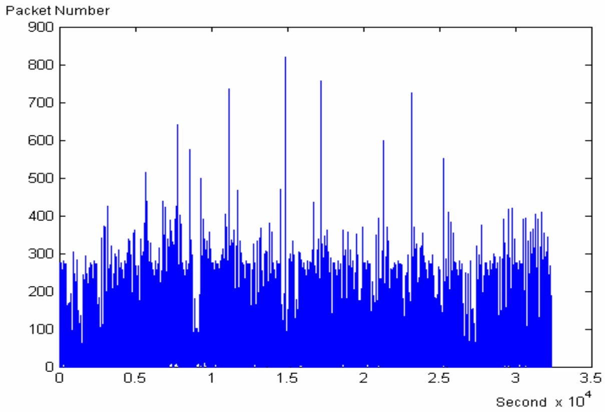


Figure 4. Merged data, 500 seconds intrusive pattern / 2000 seconds on Monday first week.

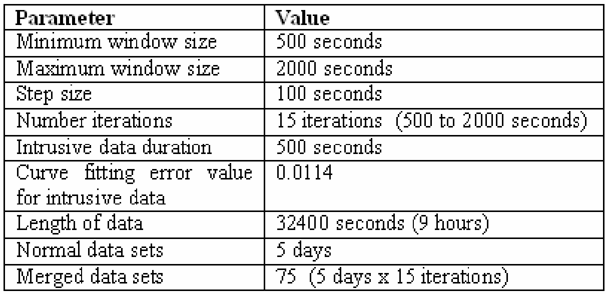
4.3. Experimental Setup

At the first stage of this experiment, the tcpdump format data can be downloaded from MIT Lincoln Lab. These network traffic data contain packet information such as flags, port numbers, ip addresses timestamp and many others. Since the size for each data set is quite huge on the average more than 300 Megabyte, it is almost impossible to analyze the data without extracting it into the database.

In this experiment we used Snort-Postgress software tool to read tcpdump format file and transfer to Postgress database in the Linux Redhat operating system. In order to query and calculate the number of network packets per second for each data set we used Perl script. Finally we converted the data to Matlab files before the data were analyzed.

To analyze the data using an iterative estimation method, it is important to determine several parameters such as in Table 1.

Table 1. Experimental parameters.



Parameter	Value
Minimum window size	500 seconds
Maximum window size	2000 seconds
Step size	100 seconds
Number iterations	15 iterations (500 to 2000 seconds)
Intrusive data duration	500 seconds
Curve fitting error value for intrusive data	0.0114
Length of data	32400 seconds (9 hours)
Normal data sets	5 days
Merged data sets	75 (5 days x 15 iterations)

The first step in our analysis was to observe the implication of increasing window size against the data insufficient probability. This includes measuring the curve-fitting error using optimization method for each window size on each data set.

Our next step was to observe the implication of increasing window size on the detection loss probability. At this stage, we used the same procedure as before with different merged data sets. From the results, we then measured the differences between both probabilities in order to get the optimum window size.

5. Results

The experiments showed the capability of iterative method to estimate the optimum window size for network anomaly detection. The estimated size will contribute to reducing the error and detection loss probability.

As shown in Figure 5, the data insufficient probabilities from Monday to Friday are exponentially reduced when the window sizes increase. The reduction indicates that the network traffic pattern is more self-similar and stationary when the window size increases. It also shows that normal network traffic data not necessarily follows the model since insufficient data have a great impact on it.

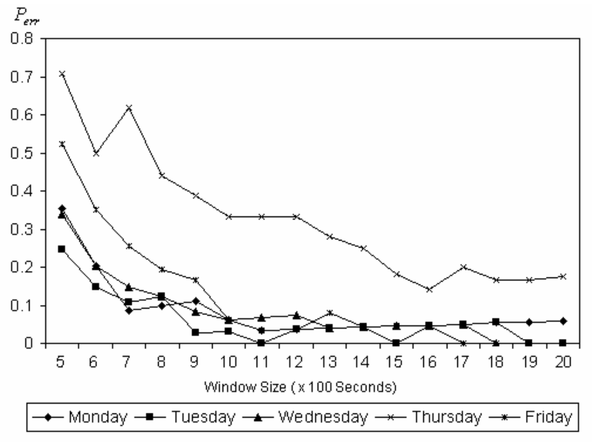


Figure 5. Data insufficient probability vs. window size.

As a result, we found that short window size will increase the false alarm where normal network traffic is considered as an intrusive pattern. However, as shown in Figure 6, by increasing window size, the detection loss probabilities will also increase. Thus, in order to balance between these two constraints, we looked at the differences between both probabilities as shown in Figure 7 and calculated the average differences as shown in Figure 8.

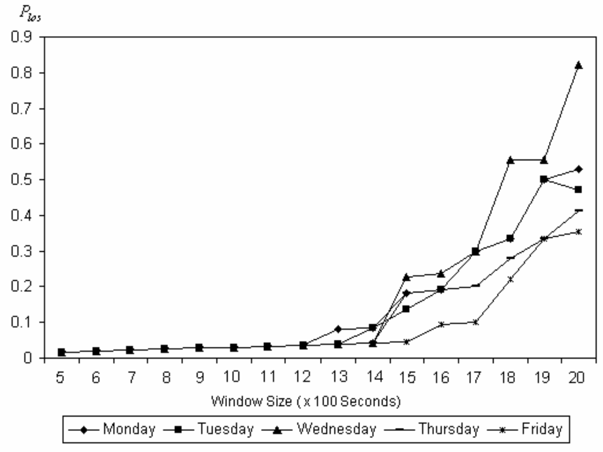


Figure 6. Detection loss probability vs. window size.

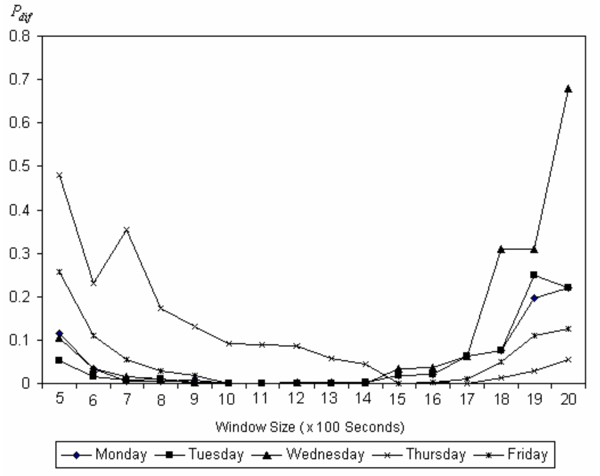


Figure 7. Probability difference vs. window size.

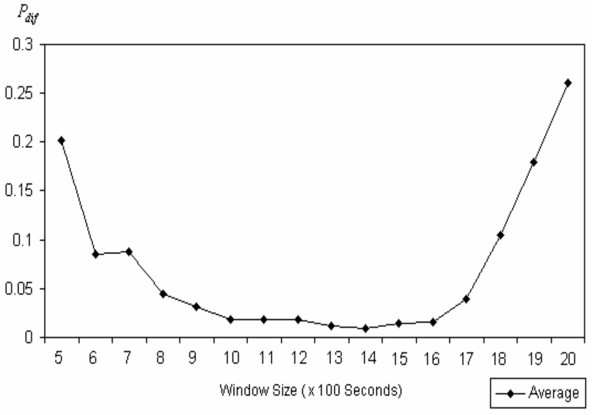
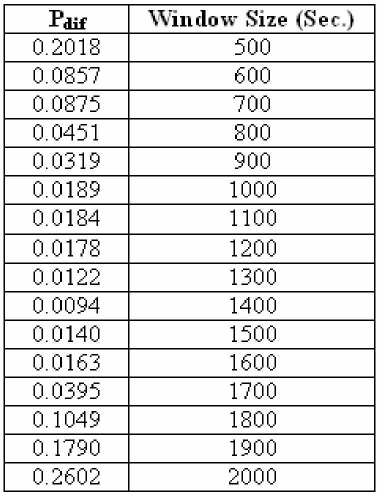


Figure 8. Average Probability Difference vs. window size

Table 2. Result.



P_{dif}	Window Size (Sec.)
0.2018	500
0.0857	600
0.0875	700
0.0451	800
0.0319	900
0.0189	1000
0.0184	1100
0.0178	1200
0.0122	1300
0.0094	1400
0.0140	1500
0.0163	1600
0.0395	1700
0.1049	1800
0.1790	1900
0.2602	2000

Figure 9 and Table 2 clearly show that the smallest probability differences at 0.0094 results from data insufficiency and detection loss probability differences. Thus in this experiment the optimum window size, ω_{opt} is 1400 seconds.

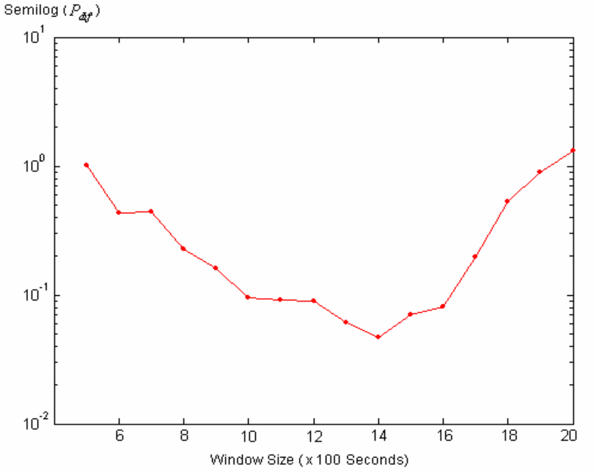


Figure 9. Log scale probability difference vs. window size.

6. Conclusion

In this paper, we have presented iterative window size estimation. The optimum window size is derived from error and detection loss probabilities. Our experiment on the MIT/DARPA data sets has shown that this method can successfully identify the optimum window size by measuring the differences of both probabilities. Thus, the selection of the window size can be justified.

We noticed that the window size is an important attribute that can increase and decrease correctness in anomaly detection. Furthermore, in order to implement this method the intrusive pattern should fulfill two conditions where it must be longer than minimum duration identified and it must be non-stationary. In this experiment the 1400 seconds window size is valid to detect abnormal patterns with the minimum duration 500 seconds with minimum curve fitting error 0.0114.

In future, we plan to enhance the method in order to adapt the normal low network traffic by using a variable window size instead of using a fixed window size.

Acknowledgements

This research was sponsored by National R&D Grant Scheme (IRPA), Malaysia.

References

[1] Houssain Kettani, “A novel approach to the Estimation of the long-range dependence parameter”, *Phd. thesis, University of Wisconsin-Madison*, 2002.

[2] Joshua W. Haines, Richard P. Lippmann, David J. Fried, Eushiuan Tran, Steve Boswell, Marc A. Zissman, "1999 DARPA Intrusion Detection System Evaluation: Design and Procedures", *MIT Lincoln Laboratory Technical Report*, 2001.

[3] Houssain Kettani, John A.Gubner, “A novel approach to the estimation of the hurst parameter in self-similar traffic”, *Proceedings conference on local computer network*, Nov. 2002.

[4] Schleifer, W., Mannle, M., **“**Online error detection through observation of traffic self-similarity”, *Communications, IEE Proceedings*, vol. 148 , no. 1, pp. 38-42 Feb.2001.

[5] Allen W.H, Marin, G.A., “On the self-similarity of synthetic traffic for the evaluation of intrusion detection systems”, *Proceedings symposium on applications and the internet*, pp. 242 – 248, Jan. 2003.

[6] M. Li, W. Jia, W. Zhao, “Decision analysis of network based intrusion detection systems for denial-of-service attacks”, *Proceedings Conferences on ICII*, vol. 5 , pp. 1-6, Nov. 2001.

[7] Nash D., Ragsdale D.J, “Simulation of self-similarity in network utilization patterns as a precursor to automated testing of intrusion detection systems”, *IEEE Transactions on*

systems, man and cybernetics, vol. 31, no.4, pp. 327 – 331, July 2001.

[8] Crovella, M.E.; Bestavros, A, “Self-similarity in World Wide Web traffic: Evidence and possible causes”. *IEEE/ACM Transactions on networking*, vol.5 no.6, pp. 835-845, December 1997.

[9] Leland, W.E., Taqqu, M.S., Willinger, W., Wilson,D.V.,“On the self-similar nature of Ethernet traffic”,*IEEE/ACM Transactions on Networking*, vol. 2, no. 1, pp. 1-15,February 1994.

[10] Paul Barford et. al. . “A signal analysis of network traffic anomalies”, *ACM Proceedings on internet measurement, SIGCOMM*, 2002.

[11] P. Abry and D. Veitch, “Wavelet analysis of long-range dependent traffic,” *IEEE Transactions on Information Theory*, vol. 44, no. 1, 1998.

[12] M. Bykova et. al., “ Detecting network intrusions via a statistical analysis of network packet characteristics”. *Proceedings of the 33rd southeastern symposium on system theory*, 2001.

[13] Willinger, W., Taqqu, M.S., Sherman, R., Wilson, D.V., “Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level”, *IEEE/ACM Transactions on Networking*, vol. 5 , no. 1,pp. 71-86, Feb.1997.

[14] Denning, D.E, “An Intrusion-detection Model”, *Transaction of IEEE on Software Engineering*, SE-13 no.2, pp. 222-223,1987.

[15] Mohd Yazid, Abdul Hanan, Mohd Aizaini, “An Iterative Estimation of Data Window Size for Anomaly Detection using Self-Similar Feature in Network Traffic”, *Proceedings of the Conference on Telematics System, Services and Applications*, pp.6-11, Mei 2004.

Mohd Yazid Idris received his B.Sc. Comp. Science and M.Sc. Comp. Science from Universiti Teknologi Malaysia in 1996 and 1998 respectively. He is a Computer Science faculty member at Universiti Teknologi Malaysia. Currently, he is a Ph.D. student working on network anomaly detection.

Abdul Hanan Abdullah received his B.Sc and M.Sc degree from the University of San Francisco, California and the Ph.D. degree from Aston University in Birmingham, UK, in 1995. He is currently the Dean at the Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia. His research interest is in Information Security. Dr Abdullah is a member of ACM.

Mohd Aizaini Maarof, Ph.D. He is an Associate Professor at Faculty of Computer Science and Information System. He obtained his B.Sc (Computer Science) and M.Sc (Computer Science) from U.S.A and his Ph.D from Aston University, Birmingham, United Kingdom in the area of Information Technology (IT) Security. He is currently leading the Group on Artificial Immune System and Security (GAINS) in the faculty. He is also a head of research in the areas of Immune-base Intrusion Detection System and Biological Inspired Cryptography System.