

The Sign Tales of Hurst Parameter: A Revisit

Mazleena Salleh¹ and Ahmad Zaki Abu Bakar²

¹Department of Computer System and Communication

²Department of Information System

Faculty of Computer Science and Information System

Universiti Teknologi Malaysia

¹mazleena@utm.my, ²zaki@utm.my

Abstract

Network traffic measurement studies have shown the presence of self-similar behavior in both local area and wide area traffic traces. In this study we investigated the magnitude of the self-similarity behavior of network traffic and its Hurst measurement, H . Our study has revealed several characteristics of H based on the network traffic changes. From the finding, H can be used as traffic load indicated only if the sampling size is large. On the other hand for a smaller sampling size, H can be used to indicate the change in traffic pattern. However network traffic does not always exhibit self-similarity property. Activities such as peer-to-peer can steer the traffic to lose this property.

Key words: Hurst parameter, self-similar, autocorrelation.

1. Introduction

Several network traffic measurement studies have shown the presence of self-similar behavior in both local area and wide area traffic traces [1] – [5]. Self-similarity is the retaining of a similar behavior or appearance of a data set over space or time. In the case of computer network, self-similarity traffic demonstrates burstiness on many or all timescales. Burstiness indicates that arrival of traffic is concentrated in short time which causes buffer to overflow and thus the need for retransmission. As a result, network utilization decreased. Researchers of [6] have shown with the aid of NS-2 simulation that network burstiness is proportional to the network load. In addition, it has also been reported that denial-of-

service (DoS) attacks on computer network is bursty in nature [7].

Although no single model is been globally accepted as definitive, the Hurst parameter, H , which describes the degree of self-similarity, holds a central place in the description of self-similar traffic. Its measurement is has been applied in the provision of quality of service (QoS), capacity planning and anomaly detection. However there are several questions that arise here which are:

- i. Is today's traffic having the same measurement of self-similarity as previous work?
- ii. Is network traffic always self-similar?
- iii. What is/are factor(s) that can affect H reading?

In this study we addressed the above issues and highlighted results which have not been extensively discussed by other researchers in the related field. The findings of the research can aid others in understanding H reading and its usage in traffic management.

This paper is organized as follows. Section 2 describes several related works in the field. Section 3 presents the methodology of the work where we describe a brief review of the platform used in conducting the experiment and the description of the network environment. Section 4 discusses the results and the findings of the study and finally, the concluding remarks are presented in Section 5.

2. Related Works

In layman's term self-similar means that the traffic is bursty at all time scales and there is no natural length

of a burst. The characteristic of bursty traffic is that it has short periods of high activity intersperse with long low activity periods. In other words, burstiness describes the density of packet arrivals on the network or the abrupt fashion in which the packets arrived. The main characteristics of a bursty traffic are the burst rate, the variance in the packet arrival rate, and variation in the packet arrival density. Average packet arrival rate, the mean packet size and the protocol and application type also affect the burstiness of the traffic.

Researchers of [8] reported that the use of simple traffic descriptors such as mean rate, peak rate and maximum burst size to evaluate the queue response lead generally to an overestimation or an underestimation of the bandwidth capacity to allocate and therefore a misuse of the network resources. Therefore to complement these traffic descriptors, an analysis on the second-order characteristics such as autocorrelation function was also performed or specifically the investigation of self-similarity of network traffic. This phenomenon plays an important role in traffic modeling because traffic correlation is an important factor in packet losses due to buffer and bandwidth limitations. It has also been shown through experimental evidence that network traffic which exhibit properties of self-similarity and long range dependent (LRD) have significant impact on network performance and the larger the burst length, the poorer the performance as the likelihood of a collision increases [9].

Self-similar behavior of the network traffic is inherited from the operation of the transport protocol as well as the usage of the network applications by the users [10-11]. Nevertheless, it also has suggested that for sub-second time periods the degree of self-similarity is insignificantly small and can be well modeled by only using only Poisson process [12]. There is also an argument whereby researchers stated that the degree of self-similarity used as parsimonious measurement for analyzing the correlation structure of analyzed data can be misleading because it assumed the underlying process is self-similar [13]. They further added that this assumption is not true for the Internet since the traffic exhibits strong daily and weekly patterns and the behavior has not changed over the years. For long enough timescale the self-similarity characteristics disappears.

3. Methodology

This study comprises of two parts: (i) experimental work and (ii) real traffic case study. The first part of the study is to investigate H for different traffic load and to identify its relationship using a network simulator NS-2 of VINT project [14]. In this experiment dumbbell topology traffic configuration was constructed. It consists of TCP senders, TCP receivers and a pair of routers with a queue size of 50. The bottleneck link has the capacity of 50Mbps, a propagation delay of 25ms and a drop tail queue. To generate self-similar traffic in the simulation, several traffic sources of Pareto distribution were generated that is by multiplexing a large number of ON/OFF data that are heavy-tailed, with the rate of 1M and shape of 1.4. Traffic loads are varied and H is measured.

For real traffic data, we applied our developed Integrated IP Traffic Measurement and Characterization Platform (IPTMChaP) to capture and analysis normal traffic data from a university campus specifically the operational network of Faculty of Computer Science and Information System (FSKSMnet), Universiti Teknologi Malaysia from May 2006 to November 2006. The network consist of approximately 1000 active users which include students, lecturers and supporting staff, ten proxies and several servers over 100Mbps link capacity that covered five departments and ten computer laboratories. The traffic data contained rich network traffic mix carrying standard network services like web, mail, FTP as well as peer-to-peer application traffic. FSKSMnet was selected for the study due to the highly intensive used of the network in all aspects of the faculty's activities, teaching, learning and faculty administration.

To measure H , we employed optimization method (OM) [15]. For OM, X_i denotes the number of bits or packets seen during the i th interval. X_i is said to be second-order stationary if its mean $E(X_i)$, does not depend on i and if the autocovariance function, $E[(X_i - E(X_i))(X_j - E(X_j))]$, depend only on i and j . With their difference $k = i - j$, the formula is rewritten as

$$\gamma(k) = E[(X_{i+k} - E(X_{i+k}))(X_i - E(X_i))]$$

For $k = 0$, the variance is

$$\sigma^2 = \gamma(0) = E[(X_i - E(X_i))^2],$$

and autocorrelation $\rho(k) = \gamma(k) / \sigma^2$

For second-order stationary process, H of $0 < H < 1$ if

$$\rho(k) = \frac{1}{2} \left(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H} \right) \quad (1)$$

Since X_i is exactly second-order self-similar, we can reduce Eq. (1) to $\rho(1) = 2^{2H-1} - 1$ and solved for estimated H as follow:

$$\hat{H} = \frac{1}{2} (1 + \log_2(1 + \rho(1)))$$

OM also provides a curve-fitting criterion in term of error function such that if the resultant error is high, then the given process fails and may not be considered to follow the self-similar model. The error function should be approximately equal to zero if X_i is close to the model. Here we will use the error function threshold of 1×10^{-3} for deciding the validity of \hat{H} . In addition we also enhance OM algorithm so that is has the capability of zooming into the traffic behavior by the using sliding window technique.

4. Results and Discussions

It is known that real IP traffic rate fluctuate accordingly to time of the day because of the influence by the change of users' population, network applications and data type that traversed in the network. To investigate the changes of traffic rate experimentally, several uncorrelated self-similar traffic with different H were aggregated. This is achieved by varying the number of connections from 15, 30, 45, to 65 at different time interval of the simulation as depicted in Figure 1. The aggregated process, referred as $Sagg_N(t)$ is defined by

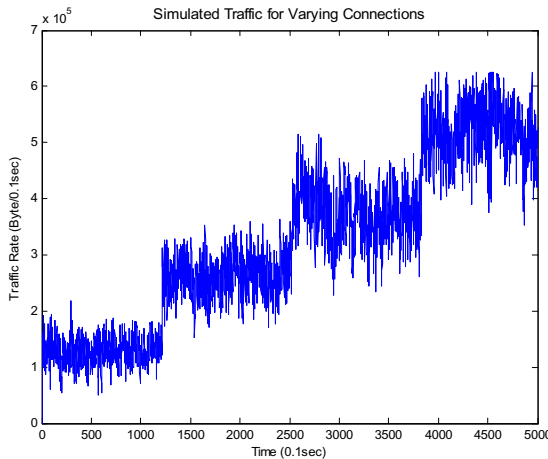


Figure 1 Aggregated Self-Similar Traffic

$$Sagg_N(t) = \sum_{i=1}^N S_i$$

where N is the number of uncorrelated self-similar processes. From the experiment, the overall or aggregated H_{agg} for $Sagg_N(t)$ is 0.98, and in fact it is the highest recorded H process as shown in Figure 2. This outcome agreed with the theorem related to aggregation of self-similar processes by [16] that stated for aggregated self-similar process comprises of N uncorrelated self-similar processes which are exactly self-similar with $H_1 = H_2 = \dots = H_N$, then $H_{agg} = H_1 = \dots = H_N$, otherwise $Sagg_N(t)$ is asymptotically self-similar with $H_{agg} = \max(H_1, H_2, \dots, H_N)$. However the varying H did not reflect the increased in traffic load. On the other hand it was observed that H decreased or increased during the transition of traffic load increment. From this exercise it can be conclude that:

- i. If there is a change in the traffic load during the sampling of H , then H_{agg} does not correlate with traffic load. In fact $H = H_{agg}$ where $H_{min} < H_{agg} \leq H_{max}$.
- ii. When there is an abrupt change in traffic load during the sampling of H , H value tend to decrease or increase. Therefore in this case, H is an indicator of an indicator of the traffic load changes.
- iii. In the event when a traffic series consist different load at different time, H will continuously varies but the direction of the change in H does not relate to the volume of the traffic load at the instant when the estimation was done.

For the second part of the work we analyzed the traffic traces of FSKSMnet. The traffic throughput of

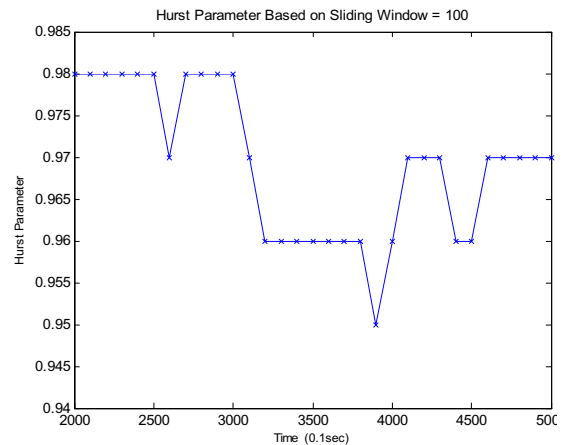


Figure 2 Estimated H for $Sagg_N(t)$

FSKSMnet is as shown in Figure 3 while Figure 4 illustrates the variation of Hurst parameter, $H_{FSKSMnet}$ for byte and packet count. For measuring $H_{FSKSMnet}$, traffic is aggregated at the time scale of 100 milliseconds with sample duration of 30 minutes. The observations are as follow:

- i. $H_{FSKSMnet}$ for packet based count is lower than the byte count.
- ii. $H_{FSKSMnet}$ is higher in the midday as compared to morning and evening reading.
- iii. For byte-count, $H_{FSKSMnet}$ is in the range of 0.785 to 0.98 with the mean reading of 0.9. This reading is higher than the previous recorded reading found in the literature survey.
- iv. Overall, $H_{FSKSMnet}$ followed traffic throughput.

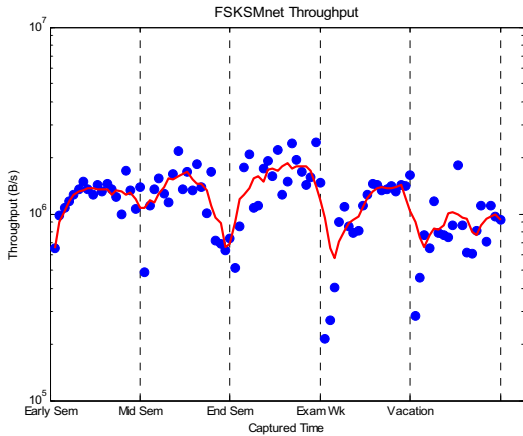


Figure 3 Throughput of FSKSMnet

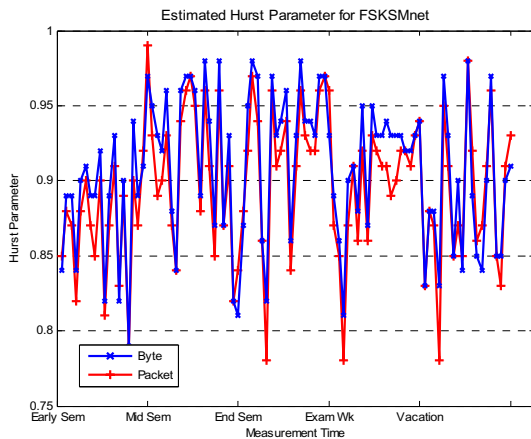


Figure 4 Estimated $H_{FSKSMnet}$

Figure 5 depicts the error function for the corresponding $H_{FSKSMnet}$ readings. It is observed that there are traces where the error exceeded the threshold value indicating the lost self-similarity property. The investigation of these traces has shown that some of them contained a high peer-to-peer activity.

For illustration purposes two traces will be discussed here: Trace I and Trace II whereby both traces contained peer-to-peer activity. The first trace is as shown in Figure 6. Figure 6 (a) is the traffic volume. It can be seen that the traffic is bursty with abrupt volume rate ranged from 3 to 5×10^6 B/s. By applying sliding window calculation of $H_{FSKSMnet}$, it was discovered that 66.77% of the sub-sampling data resulted with a high threshold error as illustrated in Figure 6 (b) and (c). In contrast to Trace I, Trace II exhibit self-similar behavior with overall $H_{FSKSMnet} = 0.99$ and the threshold error is 1.0046×10^{-4} . However it also contained peer-to-peer activities as depicted in Figure 7. Two sections of the abrupt change in traffic volume of Trace II, (a) and (b), were extracted and H was estimated for both of the sections. Section (a) traces does not follow self-similar model with $H_{FSKSMnet} = 0.8500$ and threshold error = 0.0034. This can also be seen in Figure 7 (b) and (c). On contrary, section (b) trace produced $H_{FSKSMnet} = 0.68$ and threshold error = 4.6501×10^{-4} . Even though it is self-similar, $H_{FSKSMnet}$ is relatively small as compare to any of calculated H in the traces and it is known when $H = 0.5$, it indicates a Poisson model.

From the self-similar behavior analysis of real network traffic it was observed that traffic exhibited self-similar behavior. On certain circumstances such

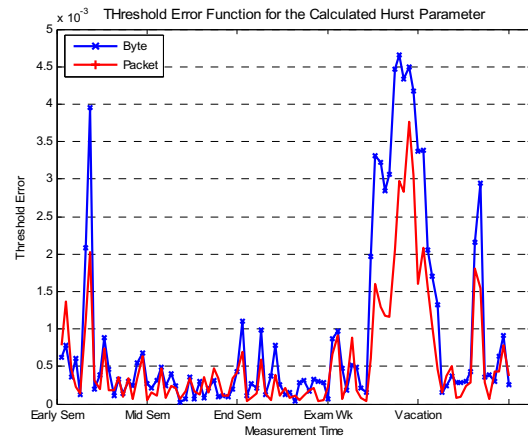
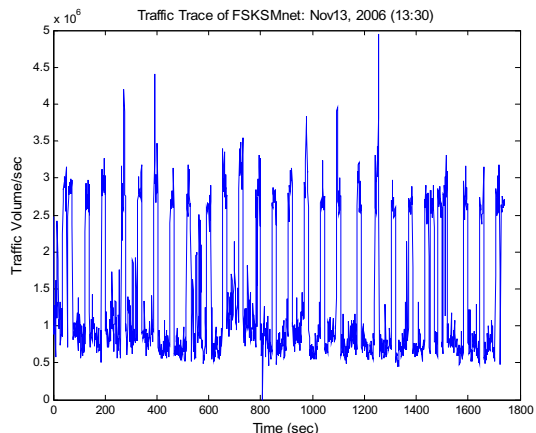
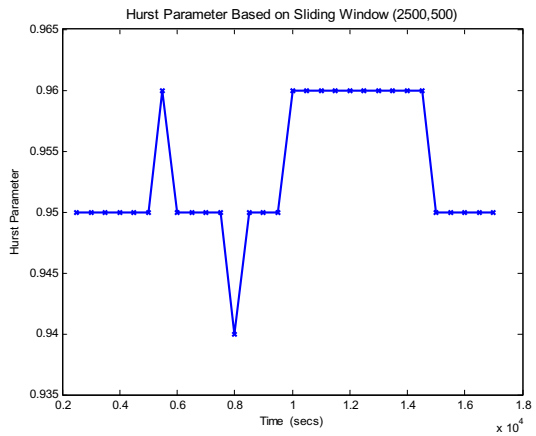


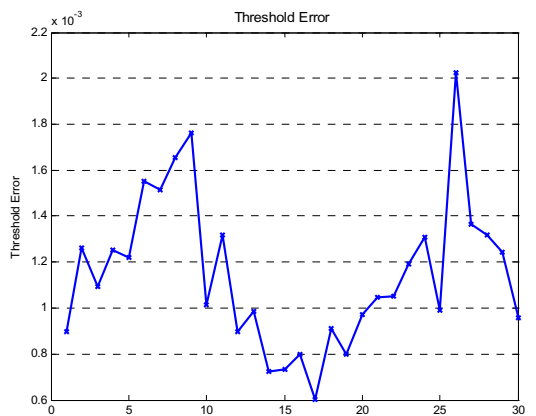
Figure 5 Threshold Error Function for $H_{FSKSMnet}$



(a) Traffic Volume



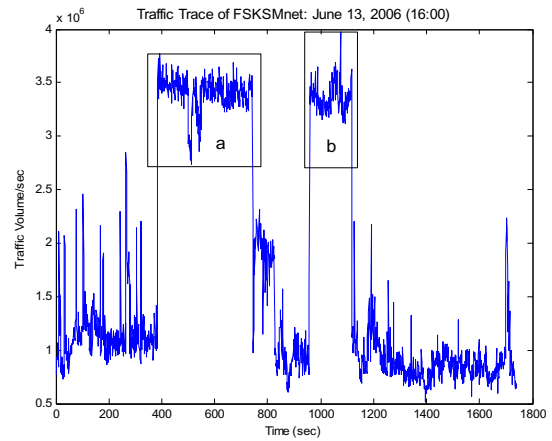
(b) $H_{FSKSMnet}$



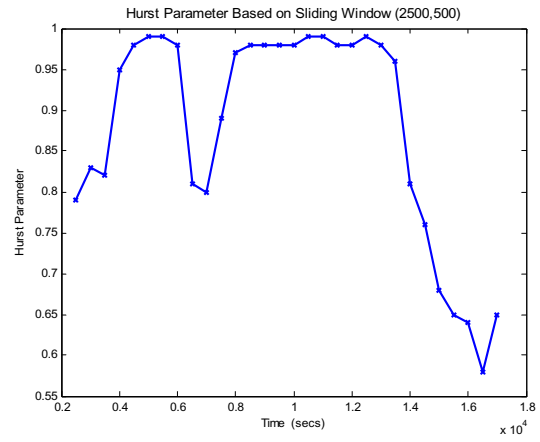
(c) Error Function

Figure 6 Trace I

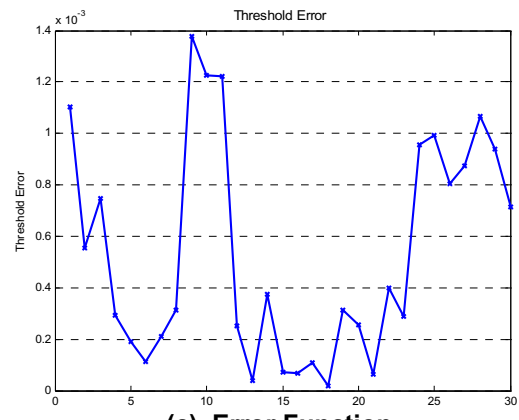
as activity of peer-to-peer, traffic volume may change and depending on the volume changes, the self-similar



(a) Traffic Volume



(b) $H_{FSKSMnet}$



(c) Error Function

Figure 7 Trace II

behavior may diminish from the trace.

5. Conclusions

In this study we have demonstrated H measurement both from simulated and real operational traffic. From the findings, H can be used to indicate traffic load if the sampling size is large. Nevertheless in zooming into the traffic traces, sub-sampling of H does not reflect the traffic load for that particular instance but instead it indicated the changes of traffic load. The real traffic exhibited a high H parameter. We believed that this is due to a higher aggregation of users and applications in the network. However, traffic does not always exhibit self-similarity property. Instead due to peer-to-peer activities the traffic may lose its self-similarity behavior. This finding has a significant impact because of the popularity and the increase of peer-to-peer applications in today's network. For our future works, we will further investigate the impact of lost of self-similarity behavior towards network performance and further characterized the used of H parameter with regard to the changes network traffic.

6. References

- [1] 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, no. 1, February 1994, pp. 1-15.
- [2] V. Paxson and S. Floyd, "Wide-area traffic: The failure of Poisson modeling", *IEEE/ACM Transactions on Networking*, vol. 3, no. 3, June 1995, pp. 226-244.
- [3] M. E. Crovella and A. Bestavros, "Self-similarity in World Wide Web traffic: Evidence and Possible Causes", *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, December 1997, pp. 835-846.
- [4] J. M. Peña, "Protocols can Make Traffic Appear Self-Similar", *Proceedings of IEEE/ACM/SCS Communication Networks and Distributed Systems Modeling and Simulation Conference*, 1997, pp. 47-52.
- [5] A. Feldmann, A. C. Gilbert, W. Willinger, and T. G. Kurtz, "The Changing Nature of Network Traffic: Scaling Phenomena", *ACM Computer Communication Review*, vol. 28, April 1998, pp. 5-29.
- [6] M. Salleh and A. Z. Abu Bakar, "Comparative Performance of TCP Variants on Self-Similar Traffic", *Proceedings of CCSP05*, Kuala Lumpur, 2005.
- [7] S. Rawat and C. S. Sastry, "Network Intrusion Detection Using Wavelet Analysis", *Proceeding of CIT 2004*, LNCS 3356, 2004, pp. 224-232.
- [8] D. Emir, E. Mohamed, and B. Ammar, "Network Performance Evaluation Using Traffic Measurements", *Control, Communications and Signal Processing, First International Symposium*, 2004, pp. 523 - 526.
- [9] A. Nogueira, P. Salvador, and R. Valadas, "Modeling Network Traffic with Multifractal Behavior", *Proceedings 10th International Conference on Telecommunications*, vol. 2, 2003, pp 1071 - 1077.
- [10] M. Crovella, and A. Bestavros, "Self-Similarity in World Wide Web Traffic, Evidence and Possible Causes", *IEEE Transaction on Communication*, 43(2), 1999, pp 1566-1579.
- [11] A. Feldmann, A. C. Gilbert, and W. Willinger, "Data Networks as Cascades: Investigating the Multifractal Nature of Internet WAN Traffic", *Proceedings of the ACM/SIGCOMM'98*, Vancouver, USA, 1998, pp 5-38.
- [12] T. Karagiannis, M. Molle, M. Faloutsos, and A. Broido, "A Nonstationary Poisson View of Internet Traffic". *Proceedings of IEEE INFOCOM*. 2004.
- [13] W. B. Gong, Y. Liu, V. Misra, and D. Towsley, "Self-Similarity and Long Range Dependence on the Internet: A Second Look at the Evidence, Origins and Implications", *Computer Networks*, vol. 48, 2005, pp 377 - 399.
- [14] L. Breslau, D. Estrin, K. Fall, S. Ford, *et al.*, "Advances in Network Simulation", *IEEE Computer*, 33(N5), 2000, pp 59-67.
- [15] H. Kettani and J. A. Gubner, "A Novel Approach to the Estimation of the Hurst Parameter in Self-Similar Traffic", *Proceedings of Local Computer Networks*, 27th Annual IEEE Conference, 2002, pp.160 - 165.
- [16] Q. Yu, Y. Mao, T. Wang, and F. Wu, "Hurst Parameter Estimation and Characteristics Analysis of Aggregate Wireless LAN Traffic", *Proceeding Wireless Networks and Emerging Technologies*, 2005.