A Fractal Modeling for Network Traffic

Sulaiman Mohd Nor
sulaiman@suria.fke.utm.my

Zuraimy Yahya
zuraimy@suria.fke.utm.my

Eko Ihsanto
eko@nadi.fke.utm.my

MiCE Department
Faculty of Electrical Engineering
Universiti Teknologi Malaysia
81310 Skudai, Johor Bahru.

Abstract. *This paper describes fractal modeling with application to broadcast traffic of switch based LAN. The traffic is represented by two variables, packet inter-arrival time and packet size. The series of these 2 variables is normalized separately and considered as a number of random points in 2-dimensional space. In the study, fractal behavior of the traffic is considered to be caused by high variability of packet inter-arrival time. Hence measuring variability of inter-arrival time is a necessity for fractal analysis. This analysis is done using long-range dependence (LRD) of variance measurement. The work compares LRD properties between real broadcast traffic and the synthetic one. The real traffic is captured from switch based-LAN, while the synthetic one is generated using Iterated Function System (IFS) method.*
Keywords: LRD, fractal, and broadcast traffic

1. Introduction

The current trend in applying multimedia applications over computer network causes a need to increase bandwidth availability to end-users. Such enhancement can be achieved by migrating from shared based network to switch based ones where dedicated end to end communication can be provided. The implementation of this new technology still allows operation of existing protocols and application including broadcast traffic for address resolution or service announcement. Although broadcast traffic reduces bandwidth availability, it is unavoidable and a necessary component for LAN to operate. By the fact that switches work like bridges, they must flood all broadcast and multicast traffic. Accumulation of the broadcast and multicast traffic from every dedicated link in the network causes broadcast radiation. Even a high capacity link can be clogged by uncontrolled huge broadcast radiation referred to as broadcast storm.

Since broadcast traffic introduces overheads on the network, there is a need to study the effect of broadcast traffic on network performance. To do this, one must have an adequate model for such traffic. A proper model can be implemented to simplify performance prediction and analysis. It is not enough to assume that, the broadcast traffic is random or bursty in nature. The phenomena and degree of randomness or burstiness should be explored, identified or measured. Although a network architecture can be known and capturing network traffic can be done to record packet size, the protocol, source and destination addresses and inter-arrival time, but still without an adequate model, it is difficult to test and measure, their influence on network performance degradation. Unfortunately, since the broadcast traffic contains accumulation of unregulated packet size and inter-arrival time, it is hard to find an adequate model for broadcast traffic representation and generation.

Chaotic behavior of Ethernet traffic has been identified since 1986. In 1991, Gusella captured traffic from 10 Mbps Ethernet LAN and concluded that the chaotic behavior is caused by variability of packet arrival in the network. For such traffic, Gusella (1991) proposed a model based on burstiness characterizations using indices of dispersion. As the model is strictly within the traditional approach of statistical analysis, it is valid only over a limited range of time scales, which is called Short Range Dependence (SRD).

A more accurate approach than that carried by Gusella was done by Leland and Wilson (1991). They presented a preliminary statistical analysis of this unique high-quality data and commented in detail on the presence of burstiness across an extremely wide range of time scales, traffic spikes ride on longer-term ripples, etc. The model of the chaotic behavior, which was identified as self-similarity or fractal behavior of aggregate Ethernet LAN traffic, is very different from previous models, such as packet-train models (Jain and Routhier, 1986), Poisson-related models (Heffes and Lucantoni, 1986) or dispersion models (Gusella, 1991).

Most known fractals are self-similar, they exhibit a similar pattern (Voss, 1988) or statistical properties (Taqqu, 1997) at different scale of magnification. If it is zoomed, in or out, a picture similar to the original one can be obtained. (Falconer, 1990).

In 1997, Willinger *et. al.* found a possibility to statistically distinguish between measured network traffic and traditional traffic models. Actual traffic exhibit correlation over wide range of time scales referred to as Long-range Dependence (LRD), while traditional traffic models typically focus on very limited range of time scales referred to as Short-range Dependence (SRD). In addition, they provided physical explanation for the observed self-similar nature of the traffic. They stated that self-similar traffic is caused by aggregation of many ON/OFF sources

0-7803-6355-8/00/ $10.00 © 2000 IEEE

whose ON or OFF-periods having high variability or infinite variance. The main themes of the study are LRD and self-similarity. The former involves the tail behavior of the autocorrelation function of a stationary time series, while the later typically refers to the scaling behavior of the finite dimensional distributions of continues time or discrete time process.

After studying fractal characteristic of the network traffic, Taqqu *et. al* .(1997) concluded that measured LAN and WAN traffic traces, with the sample means subtracted, are well modeled by random processes that are either exactly or asymptotically self-similar. They prove, that network traffic exhibit self-similar single fractal rather than multifractal behavior.

2. Fractal analysis using LRD measurement

Long Range Dependence (LRD) variance-time plot is used to obtain qualitative characterization of the correlation present in the traffic data. The plot is obtained by calculating variance of mean, which are calculated from various size of data aggregation.

A process Z is said to exhibit long range dependence (LRD) if it possesses strong correlation for large lags. LRD can be characterized in terms of the behavior of the aggregated processes

$$Z^{(m)}[n] = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} Z[i] \qquad (2.1)$$

where m is the level of aggregation, which represents the number of element in each segment of the entire process. While n is the number of samples, and k is the index number of the segment.

If H is hurst parameter of Z and

$$var(Z[n]) = m^{2-2H} var(Z^{(m)}[n]) \qquad (2.2)$$

it is considered that Z exhibit second-order self-similar process. For such process, a log-log plot of the variance of $Z^{(m)}[n]$ as a function of m is strictly linear with a slope $2-2H$ (Riedi *et. al.*, 1999). In the opposite, non-linear exhibition of the variance plot indicates multifractal behavior of the process.

3. Multifractal IFS model for Network traffic

This section contains several steps for actual implementation involved to capture the fractal properties of real traffic pattern including the IFS parameters estimation. In what follows, foundation for a basis on further work applied to networks will be described. This foundation is on extension of the work done by Barnsley (1988). The work that follows will be a contribution that was developed to solve the traffic-modeling problem as mentioned in section 1. Block diagram of work flow is seen in Figure 3.1. The procedure using the IFS model to generate the synthetic traffic is presented here. In addition, the methods for comparing the real traffic to the synthetic one are described.

Data Collection processs as seen in Figure 3.1. is done by sniffing the network segment and capturing the broadcast traffic packets. Four series of the real traffic data were captured from a LAN domain identified as subnet 118 in Faculty of Electrical Engineering, Universiti Teknologi Malaysia.

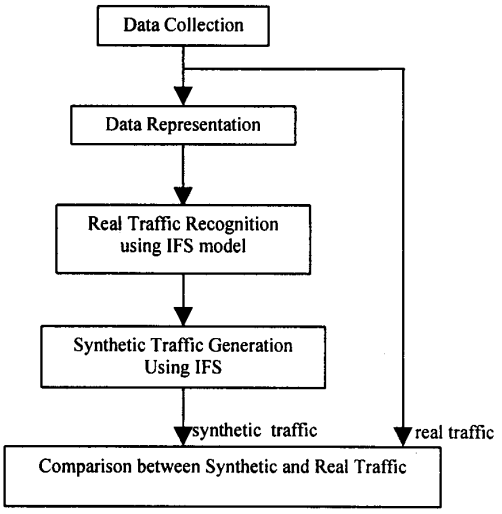

Figure 3.1 Work flow

The number of packets in each of the four files were 62185; 72641; 52074 and 18672. Capturing was done to within accuracy of 1 μs. Beside the amount of packet and time precision, there are three reasons to assume that data is sufficient. These are the amount of protocols that exist, the amount of station sending the broadcast traffic and the time duration of capturing packet. The three protocols existed, i.e. IP (TCP, IP, UDP and ICMP), IPX and NetBeui. The packets were generated by about 200 stations for a duration of 60 hours 58 minutes.

From the several fields of information captured by the sniffer, only two fields will be analyzed, i.e. inter-arrival time and packet size fields. To allow analyzing these fields by another software tool, two steps of converting process must be done. First, the captured data is printed out to text file. Second, the text file obtained by the first step is converted to dbf format using either Microsoft Excel or Microsoft Access. In this step, all fields except inter-arrival and packet size fields are removed. Later, the dbf file obtained by second step will be read by a custom software in C++.

The next step after capturing data is data representation as seen in Figure 3.1. Before being analyzed, data consisting of series of inter-arrival time and packet size must be represented in one or two variable series. To simplify the analysis, the series are normalized. Although the inter-arrival time and the packet size have different units, they can be represented to time unit, i.e., ON duration for the inter-arrival time representation and OFF duration for the packet size representation. OFF duration is obtained by multiplying the packet size with duration of one byte data transfer.

Due to a large difference between ON and OFF duration in the broadcast traffic, it is not practical to represent them in the same unit. ON varies from 6 ms to 150 ms, while OFF varies from 0 to 10 second with 1 μs precision. To simplify data analysis and synthesis, these two variables were normalized separately. Minimum and maximum values for ON variables are 60 and 1514, respectively, while for OFF variables are 0 and 10 million, respectively.

The series can be considered as point series in 2 dimensional space $(x,y \in R)$, where x represents the normalized inter-arrival time and y represent the normalized packet size. By the fact that certain protocols have certain packet size and their inter-arrival behavior, it is considered that both x and y are inter-related. The example of this representation can be seen in Table 3.1.

n	Source Data		Normalized data	
	Inter-arrival (μs)	Packet size (bytes)	x_n	y_n
0	0	248	0	0.129298
1	4737594	180	0.473759	0.082531
2	14942	98	0.001494	0.026135
3	18652	546	0.001865	0.33425
4	1837	546	0.000184	0.33425
5	2897	546	0.00029	0.33425
6	2589	546	0.000259	0.33425

Table 3.1. An example of data normalization

The other kind of representation was implemented by converting two variable series into one variable series based on the nature of source data. Since source data consists of inter-arrival time and packet size series, they can be represented as a series of single type of variable, i.e. bytes aggregation per unit time. Hence, bytes aggregation is a function of inter-arrival time and packet size.

Since the real traffic is assumed as fractal pattern, to recognize it, an appropriate model for fractal pattern recognition must be used. In this work, recognizing the real traffic pattern is done by using IFS model. As seen in Figure 3.1, this process takes place after data representation step.

The series that will be treated in what follows has chaotic behavior, an anomaly that has been explained in section 1. Since the series has chaotic behavior, it is hard to analyze it using traditional mathematics. One of the mathematical approach to model this chaotic behavior is by assuming it as a number of points generated by iteration process called Iterated Function System (IFS) (Barnsley, 1988).

After several numerical trials were done, a simple method used to estimate optimal parameters of IFS called the estimation procedure for best-fit LTFs was proposed. This estimation method was implemented to recognize the real traffic patterns. After IFS parameters of the real traffic pattern were found, the parameters were used to generate the synthetic traffic series. This IFS traffic generator consists of five Linear Transformation Functions, single uniform random generator and probability distributions for several time segments. The synthetic traffic pattern was generated using Random Iteration Algorithm.

4. Result and Discussion

After obtaining synthetic traffic, it should be done certain analysis to compare the synthetic traffic with the real one. The comparison method used in this work is LRD of variance measurement. This method is applied to explore high variability or heavy tailed distribution of packet size and its inter-arrival time.

LRD analysis uses variance time plot at various scales to estimate LRD of one-dimensional fractal pattern. This approach is done to obtain a qualitative characterization of the correlation present in the data. Figure 4.1 through 4.6 shows the results of LRD and variance measurement. All graphs on the left side are comparison results between the synthetic traffic and the real traffic that was captured from 10:21 p.m. on July 27 1999 until 1:08 p.m. on July, 28 1999. On the right side, the real traffic was captured on July 29 1999 from 1:22 p.m. until 8:17 p.m. All traffic traces consist of 32769 packets.

Figure 4.1 shows that both synthetic and real traffic curves exhibit hyperbolic shape. It indicates slow decay of autocorrelation. The figure also shows that for the trace 1 of real data, the synthetic trace exhibit faster decay than the real one (left side), while for trace 2 of real data, it is shown the contrary fact (right side).

As stated by Riedi *et. al.* (1999), by the fact all curves shown in figure 4.1 and 4.2 have multiple slope, it is concluded that the broadcast traffic captured from switch based LAN exhibit self similar multifractal behavior rather than single fractal. This fractal behavior is exhibited by both packet size and interarrival series.

On Figure 4.2, it can be seen that the real traffic curve has a unique shape that fails to be imitated. Although packet size is random, they are constrained to limit variation of size, in addition, sometimes the same packet size are generated consecutively. This characteristic is quite different compared to interarrival series characteristic. The interarrival series is really random with unlimited variation.

Figure 4.3 and 4.4 show different views of LRD measurement results. The single variable series being analized, i.e. interval series and bytes aggregation series are calculated as a function of packet size and its interarrival time. The interval series was obtained by calculating time interval between consecutive 200 bytes of the traffic. The bytes aggregation series was obtained by aggregating bytes of the traffic every 2 second.

As can be seen in Figure 4.3, although both synthetic and real curve exhibits LRD, the synthetic curve is too smooth compare to the real one. It is assumed that defect of real traffic curve is caused by unique behavior of packet size series.

Figure 4.4 shows very different result between graph on the left and right side. The same case with Figure 4.3, these bad graphs are assumed caused by improper

choice of byte size for figure 4.3 and interval for figure 4.4. As can be seen in figure 4.5 and 4.6, variation of byte size and interval obtains similar view of real and synthetic traffic.

5. Conclusion

From the work which was carried out, it was shown that broadcast traffic exhibits self similar multifractal behavior. The finding from this work can be used to generate simulated traffic pattern for network test bed or simulation package.

References

Barnsley, M. (1988). "Fractals Everywhere". New York: Academic Press, Inc.

Falconer, K.J. (1990). "Fractal Geometry: Mathematical Foundations and Applications." New York: John Wiley and Sons.

Gusella, R. (1991). "Characterizing the Variability of Arrival Processes with Indexes of Dispersion". *IEEE Journal on Selected Areas in Communications*, Vol. 9, pp. 203-211.

Heffes, H. and Lucantoni, D. M. (1986). "A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance", *IEEE Journal on Selected Areas in Communications* Vol. 4, 856-868.

Jain, R. and Routhier, S. A. (1986). "Packet Trains: Measurements and a New Model for Computer Network Traffic", *IEEE Journal on Selected Areas in Communications*, Vol. 4, 986-995.

Leland, W. E. and Wilson D. V. (1991). "High Time-Resolution Measurement and Analysis of LAN Traffic: Implications for LAN interconnection", Proceedings of the IEEE INFOCOM'91, Bal Harbour, FL, 1360-1366.

Riedi, R.H., Crouse, M.S., Ribeiro, V.J., and Baraniuk, R.G. (1999). "Multifractal Wavelet Model with Application to Network Traffic." *IEEE Transactions on Information Theory* (Special Issue on Multiscale Signal Analysis and Modeling), Vol 45. 992-1018.

Taqqu, M.S., Teverovsky, V., and Willinger, W. (1997). "Is network traffic self-similar or multifractal?" *Fractals* 5, pp. 63-73.

Voss, R.F. (1988). "Fractals in nature: From characterization to simulation." in Peitgen, H.O. and Saupe, D. (Eds.). "The Science of Fractal Images." New York: Springer-Verlag. 21-70.

Willinger, W., Taqqu, M.S., Sherman, R. and Wilson, D.V. (1997). "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.1-86.

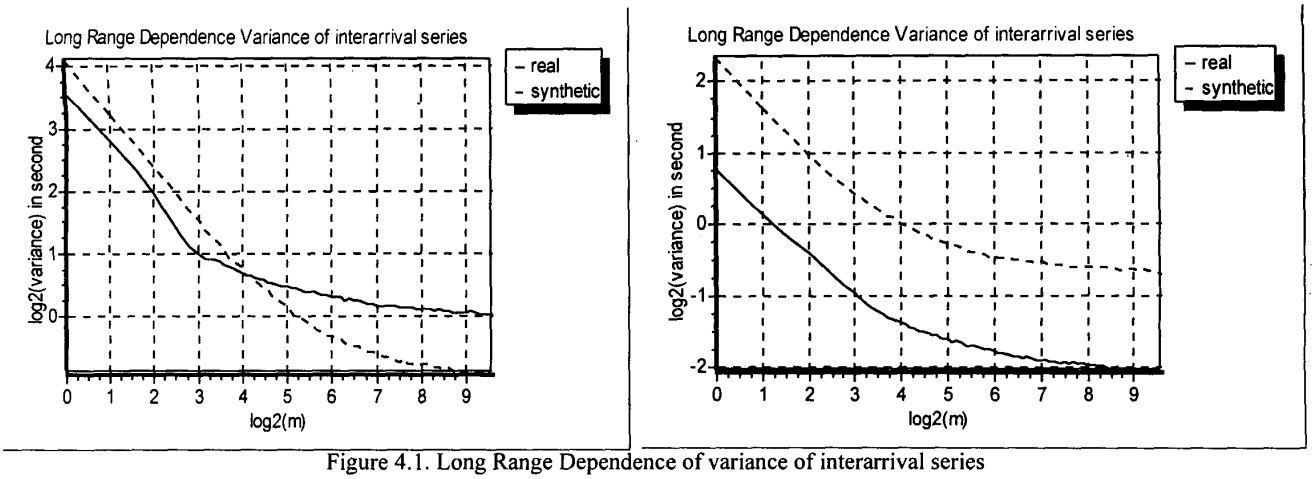

Figure 4.1. Long Range Dependence of variance of interarrival series

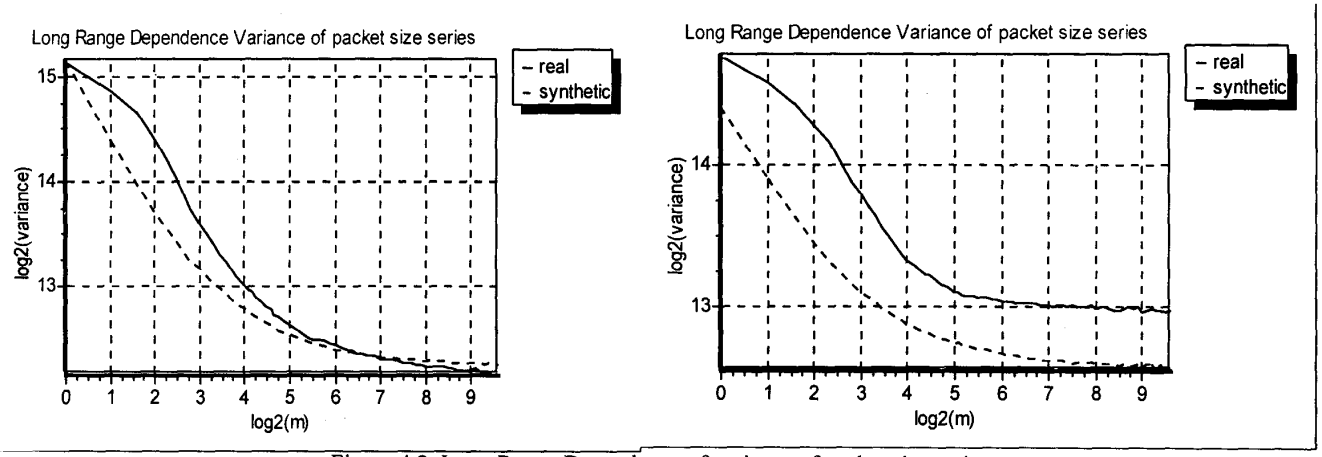

Figure 4.2. Long Range Dependence of variance of packet size series

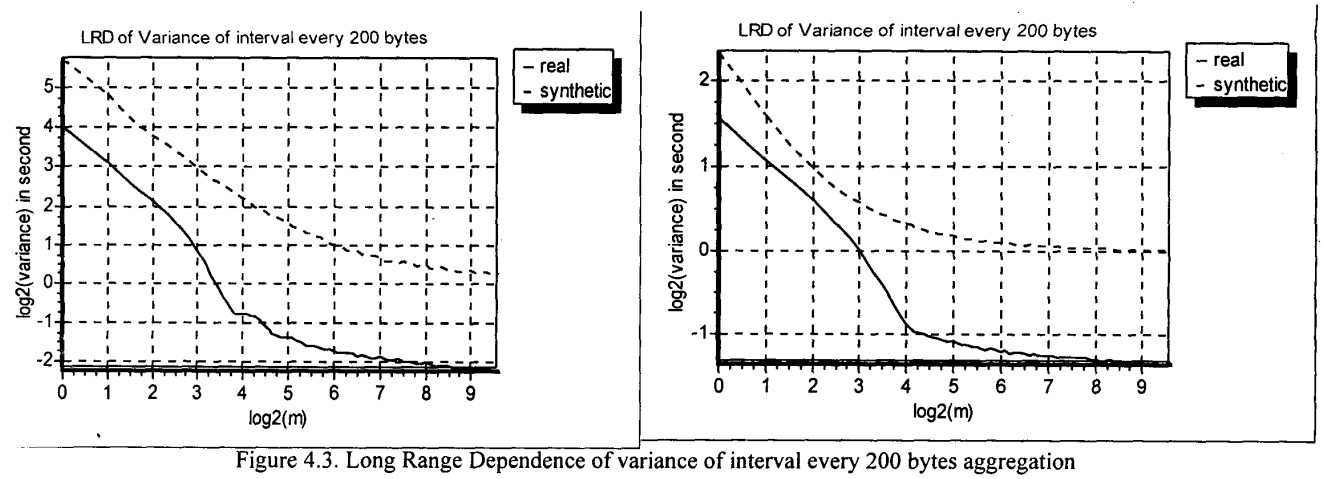

Figure 4.3. Long Range Dependence of variance of interval every 200 bytes aggregation

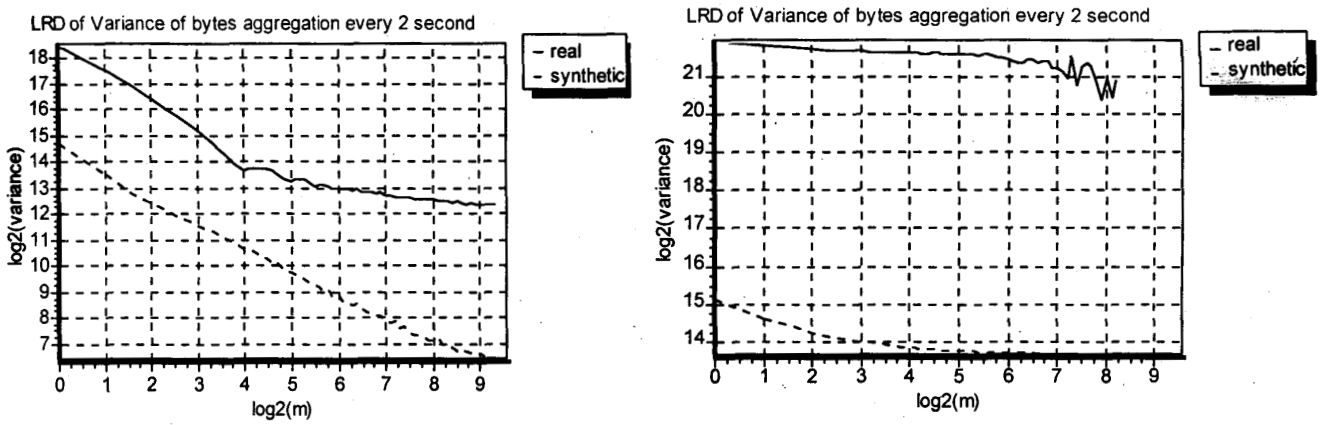

Figure 4.4. Long Range Dependence of variance of bytes aggregation every 2 second

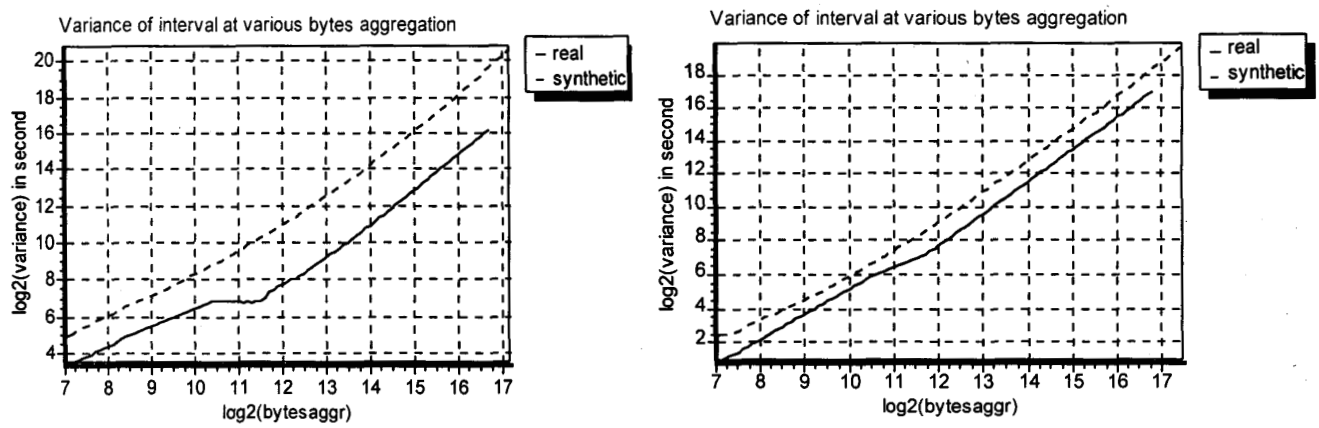

Figure 4.5. Variance of interval at various bytes aggregation

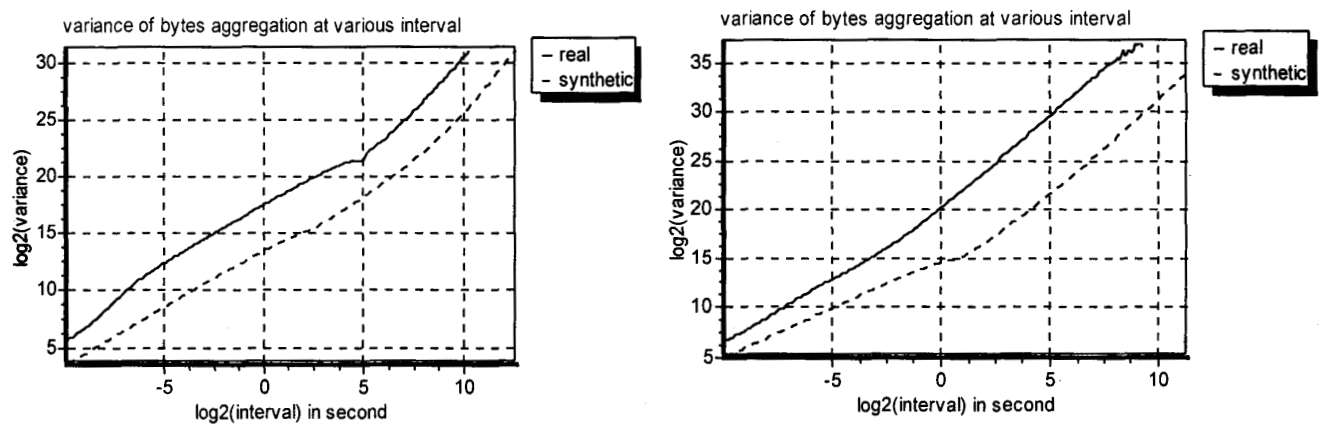

Figure 4.6. Variance of bytes aggregation at various interval