

July 2011

Mathematics in Internet Traffic Data Analysis

Sasmita Acharya

Dept. of Computer Sc. & Engg., IGIT, Sarang, Dhenkanal, Orissa, talktosas@gmail.com

Sasmita Mishra

*Asst. Prof., Dept. of Computer Sc. & Engg., IGIT, Sarang, Dhenkanal, Orissa,
sasmita.mishra.csea@gmail.com*

S.N. Mishra

*Professor & Head, Dept. of Computer Sc. & Engg., IGIT, Sarang, Dhenkanal, Orissa,
sarose.mishra@gmail.com*

Follow this and additional works at: <https://www.interscience.in/ijcsi>



Part of the [Computer Engineering Commons](#), [Information Security Commons](#), and the [Systems and Communications Commons](#)

Recommended Citation

Acharya, Sasmita; Mishra, Sasmita; and Mishra, S.N. (2011) "Mathematics in Internet Traffic Data Analysis," *International Journal of Computer Science and Informatics*: Vol. 1 : Iss. 1 , Article 12.

DOI: 10.47893/IJCSI.2011.1011

Available at: <https://www.interscience.in/ijcsi/vol1/iss1/12>

This Article is brought to you for free and open access by the Interscience Journals at Interscience Research Network. It has been accepted for inclusion in International Journal of Computer Science and Informatics by an authorized editor of Interscience Research Network. For more information, please contact sritampatnaik@gmail.com.

Mathematics in Internet Traffic Data Analysis

Sasmita Acharya

Lecturer, Dept. of Computer Sc. &
Engg., IGIT, Sarang,
Dhenkanal, Orissa
Email:talktosas@gmail.com
Ph:06760265776

Sasmita Mishra

Asst. Prof., Dept. of Computer Sc. &
Engg., IGIT, Sarang,
Dhenkanal, Orissa
Email:sasmita.mishra.csea@gmail.com
Ph:06768267056

S.N.Mishra

Professor & Head,
Dept. of Computer Sc. & Engg.,
IGIT, Sarang,
Dhenkanal, Orissa
Email: sarose.mishra@gmail.com
Ph:06768267056

Abstract

The Internet traffic data have been found to possess extreme variability and bursty structures in a wide range of time-scales, so that there is no definite duration of busy or silent periods. But there is a self-similarity for which it is possible to characterize the data. The self-similar nature was first proposed by Leland et al [1] and subsequently established by others in a flood of research works on the subject [2]-[5]. It was then a new concept against the long believed idea of Poisson traffic. The traditional Poisson model, a short ranged process, assumed the variation of data flow to be finite but the observations on Internet traffic proved otherwise. It is this large variance that leads to the self-similar nature of the data almost at all scales of resolution. Such a feature is always associated with a fractal structure of the data. The fractal characteristics can exist both in temporal and spatial scales. This was indicated by Willinger and Paxson [6], as due to the extreme variability and long range dependence in the process. Presently, one of the main research interests in the field of Internet traffic is that of prediction of data which will help a network manager to render a satisfactory quality of service. Before preparing a model of prediction, one of the important tasks is to determine its statistics. Any model to predict the future values will have to preserve these characteristics.

Keywords: Mathematics, Internet traffic, fractal

1.0 INTRODUCTION

Traffic models are important to network performance analysis. They are required to capture the statistical characteristics of real traffic efficiently and accurately. The development of traffic modeling relies on the advances of traffic analysis. One of the traffic models is derived from an On-Off queuing model. In this model, “On” represents a busy data transmission period and “Off” represents silence with no data transmission. Statistically, if “On-Off” periods are heavy-tailed, aggregating a large number of these processes will result in long-range dependence [5]. On the

other hand, self-similarity of a process is a property represented using its power spectrum. By exploiting the power spectrum, a number of traffic models [7]–[10] have been proposed since the late 1990s and early 2000s. In particular, the Haar wavelet has been used to analyze and to model network traffic in recent years. Among a number of network traffic models, fractional autoregressive integrated moving average (FARIMA), $M/G/\infty$ & Wavelet traffic models are three famous examples. This paper mainly discusses the concept of mathematics and its application in the study of statistical characteristics of internet data.

2.0 TCP TRAFFIC

In this section, an experimental study of the fractal traffic control mechanism is presented using the concept of efficient bandwidth on TCP traffic. Normally, Web, email, and FTP sessions are carried by TCP traffic. These are not time-critical applications.

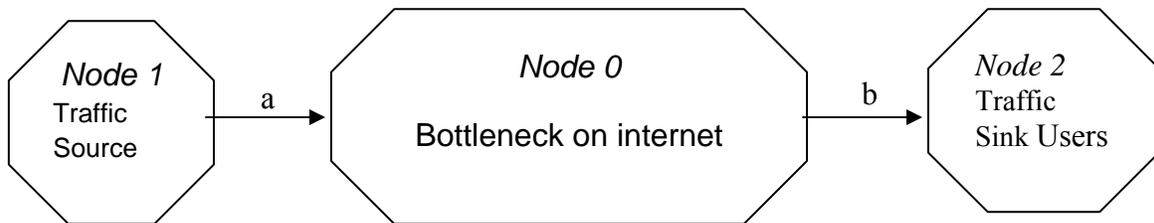


Figure 1: The General Network Model

Choosing a timescale of 100ms or even a second should not affect services. But, packet bursts originate at millisecond or sub-millisecond timescales. Removing traffic burstiness will greatly increase the network efficiency. In this experiment a large number of TCP flows simulated in NS-2 originate from “node 1” to “node 2”. In the middle is a bottleneck “node 0”. This simplified network model emulates a typical Internet traffic scenario and is shown in Figure 1. Here, “node 1” represents an Internet server domain, which typically has a high bandwidth for downstream data transfers; “node 2” represents a user group, which receives packets from the server group. “Node 0” acts as an Internet bottleneck, at which traffic experiences limited bandwidth, buffer delay and packet loss.

2.1 An Experiment

In this experiment, four tests were conducted with the controlled bandwidth at link (b) of Figure 1 at 0.3, 0.5, 1, 10Mb/s. Table 1.0 lists the experiment parameters.

Items	Contents
Traffic Source	<ul style="list-style-type: none"> - Poisson connections with Pareto distributed size - Number of connections: 3000 - Connection arrival time: 0.6s - Connection Pareto size: 12 - Connection Pareto shape: 1.5 - Traffic type: TCP
Links and Bandwidth	<ul style="list-style-type: none"> - Link (a): 0.3, 0.5, 1, 10 (Mb/s) - Link (b): 1Mb/s, the bottleneck link with one-way time 75ms and buffer capacity of 10 packets.

Table 1.0 Experiment parameters

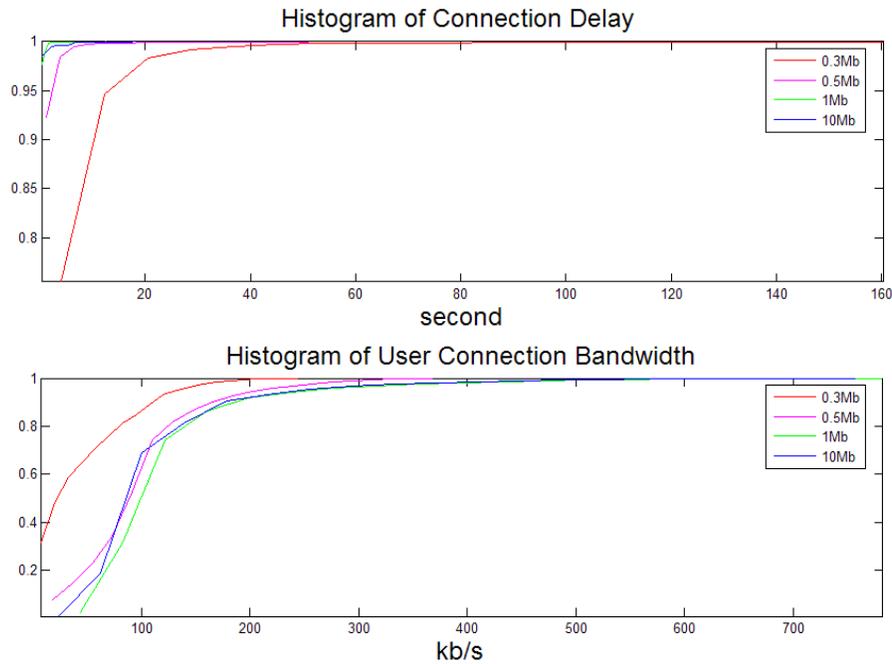


Figure 2: Experiment results (1)

Figure 2 shows histograms for delay and bandwidth. From these histograms, it is found that the case of link “a” with 1Mb/s is the fastest one among the four cases. The 10Mb/s link is a little slower because it has no bandwidth control and has packet drops at the bottleneck. The case with only 0.5Mb/s performs very well and has moderate throughput, but for the 0.3Mb/s link the performance is greatly degraded. The benchmark test has a 10Mb/s bandwidth without the bottleneck. Compared to the benchmark trace, the maximum connection delays are 164.356s (0.3Mb), 52.504s (0.5Mb), 12.986s (1Mb), and 18.536s (10Mb), respectively. In Figure 2, for the average of connection time factors, both the 1Mb and 10Mb cases are close to 1, the 0.5Mb is just over 2, and the 0.3Mb is greater than 9. From those numbers, it is obvious that the 0.3Mb link is congested, the bandwidth at 10Mb/s is too fast for the bottleneck, and 1Mb/s is sufficient enough but may not be efficient (this is shown by the difference on throughputs between the source’s and the user’s in Figure 3 upper right).

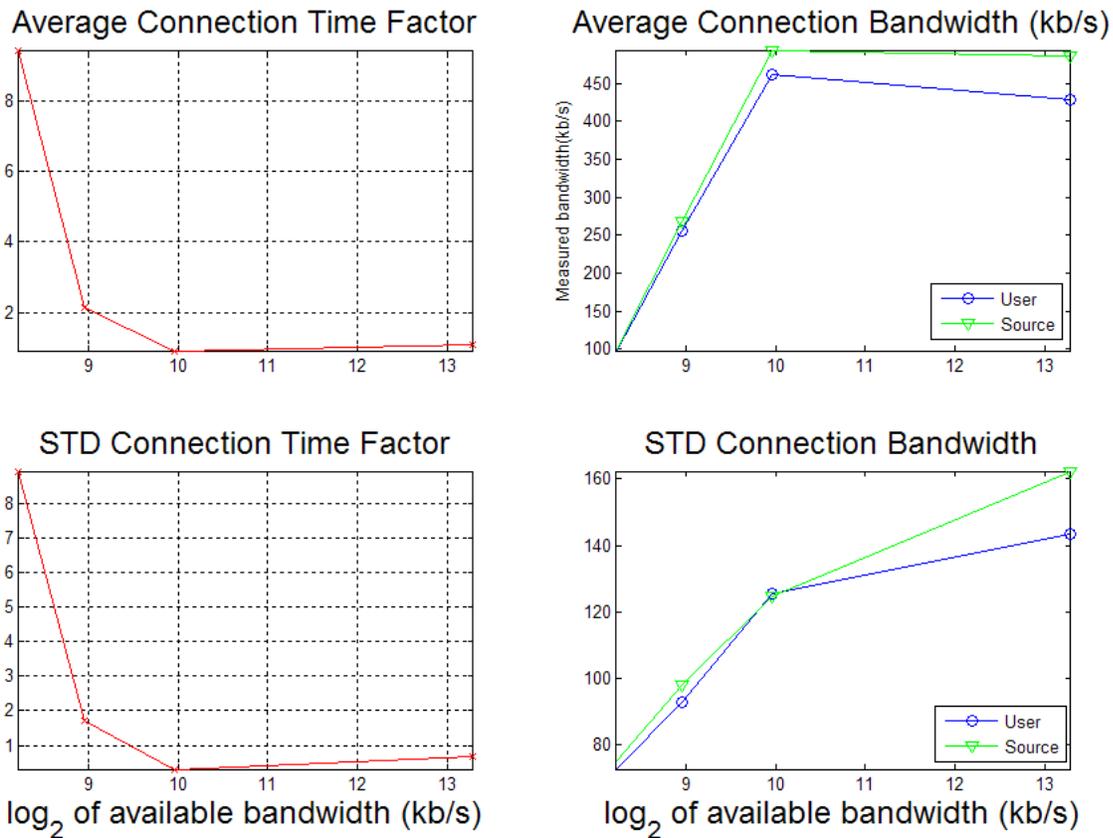


Figure 3: Experiment results (2)

Figure 4 plots the efficient bandwidth. For timescales less than 1s, the 0.3Mb link reaches the maximum of 0.3Mb/s so traffic is congested below this timescale. The 0.5Mb link has no problem under the 1s timescale. Its peak bandwidth is about 400kb/s at the smallest timescale of 10ms. Comparably, both the 1Mb link and the 10Mb link are close to 480kb/s at the maximum. Increasing bandwidth from 0.5Mb/s to 1Mb/s only has a small gain in throughput (400kb/s to 480kb/s). So, it is implied that the 0.5Mb link is the best among the four cases. However, in reality, choosing the right bandwidth between 0.3Mb/s to 1Mb/s is not tightly defined. The best value also depends on the network environment, i.e., network traffic loads from other traffic flows. Using efficient bandwidth gives an accurate estimation on how to multiplex fractal traffic flows to share the bandwidth at a common link.

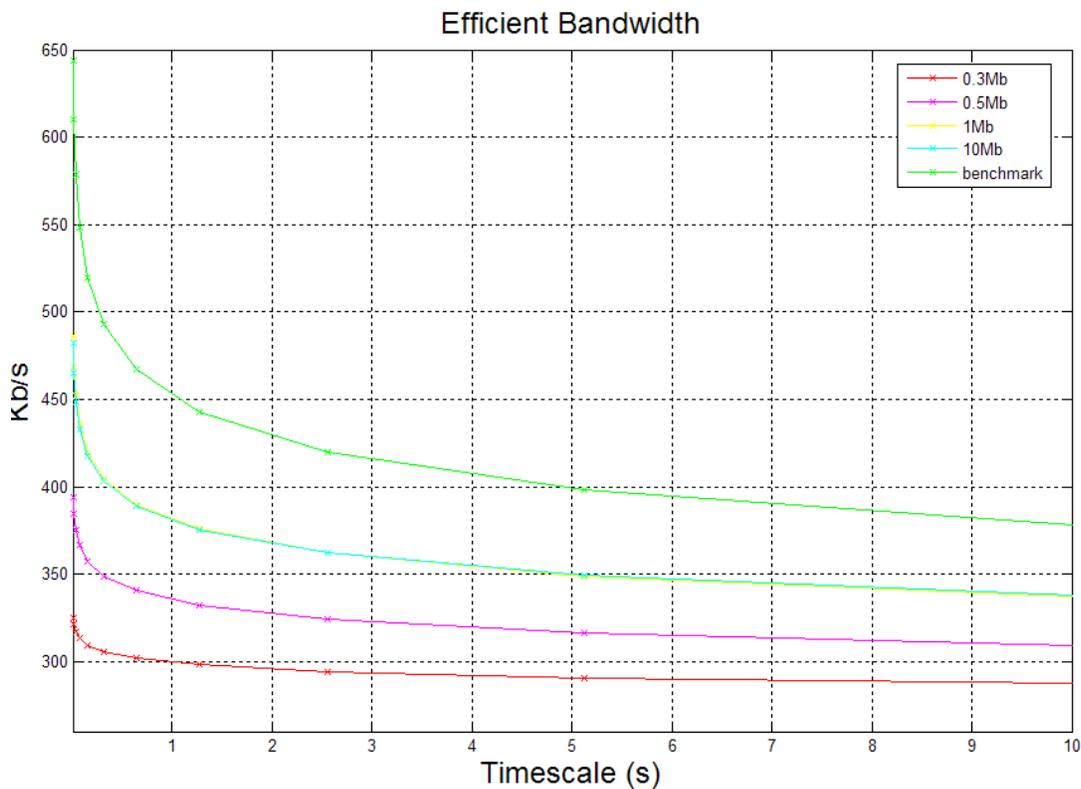


Figure 4: The efficient bandwidth

Figure 5 gives the detail of the connection time factor and the packet loss. The outline of the connection time factor in the 0.3Mb/s case (red) resembles a fractal image (Von Koch curve) in a random fashion, and that in the 0.5Mb/s case also looks similar but in a smaller scale. This demonstrates the traffic fractal behaviors in another aspect: when fractal traffic is constrained (by bandwidth), network delays increase nonlinearly. This finding invalidates the formula: $bandwidth = traffic / time$. The figure shows delay will increase exponentially with decreasing bandwidth. Figure 5 zooms in to show the variations in the connection time factor in finer detail, and the above property remains similar. Furthermore, in Figure 5 bottom plot, it is noticed that the 10Mb link (blue dots) has packet losses in some connections. Large connection time factors in the 10 Mb link are caused by their packet losses. The 10Mb case is the only one suffering packet losses in this experiment. It is apparent that without bandwidth control traffic is very bursty at the packet level, and those packet losses in fractal traffic occur in bursts which are non-uniform.

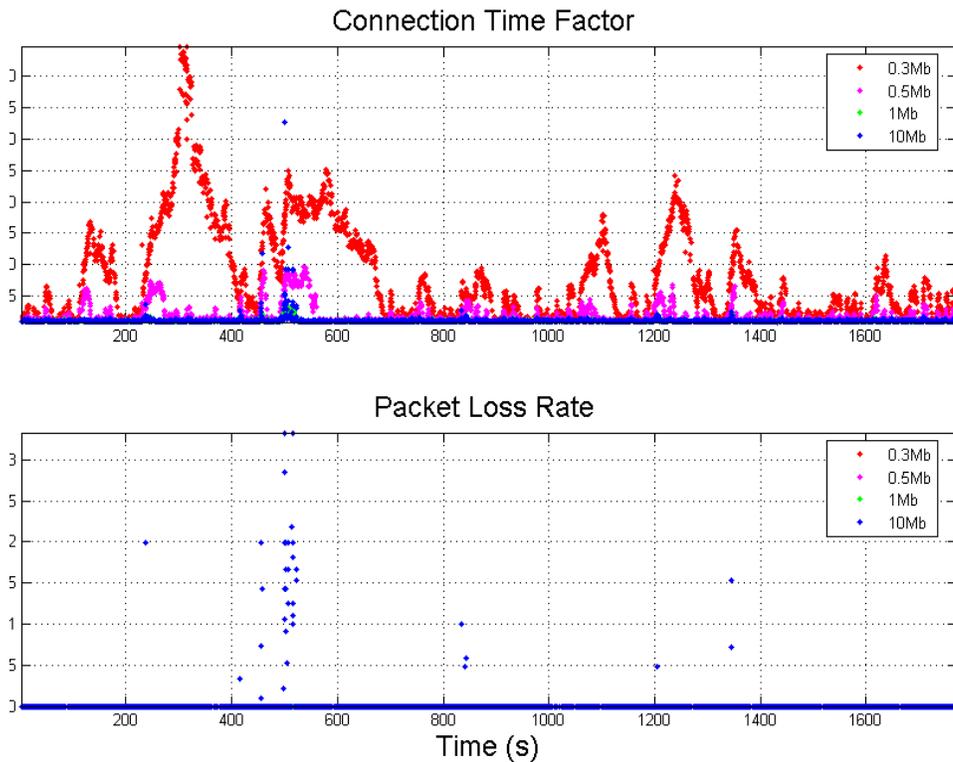


Figure 5: Connection Time Factor in smaller time ranges (continued..)

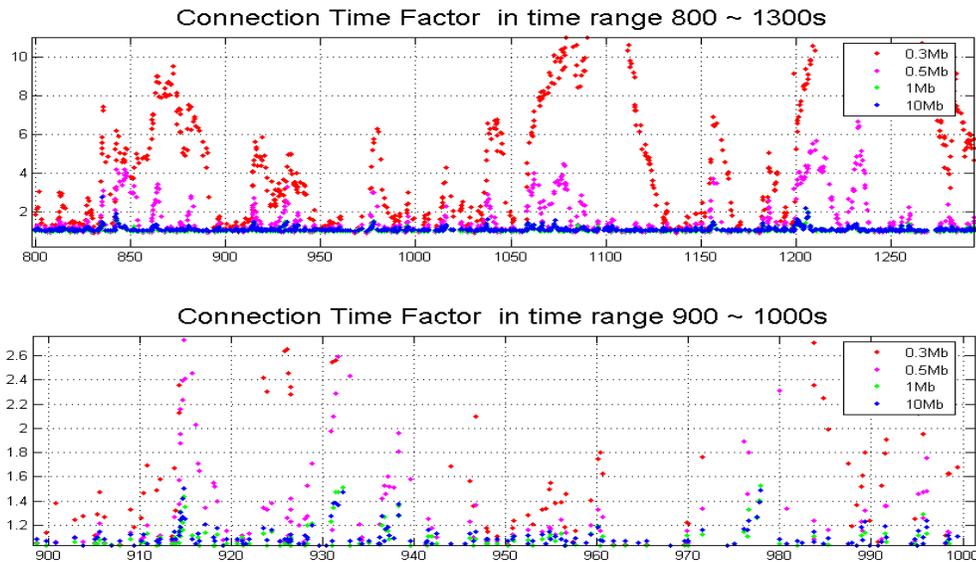


Figure 5: Connection Time Factor in smaller time ranges

3.0 DISCUSSION

In network engineering, efficiency is a key objective in network provisioning. Using EB the network performance can be improved in the following aspects:

- (1) It reduces the degree of traffic burstiness, and more fractal traffic flows can be multiplexed without interference. Thus, the network efficiency can be improved.
- (2) EB is more accurate in measuring a fractal traffic flow bandwidth than the other conventional methods, such as the mean. Because of the scaling of fractal traffic, a flow with a small mean could have a large EB at a small time scale. The scaling exponent plays a very important part in fractal traffic.
- (3) It is easy to calculate EB and to measure it in real-time to provide adaptive control.

In addition, the following aspects of EB are observed:

- (1) As $\alpha \rightarrow 1$, EB increases slowly with decreasing timescale. So, for a given time scale, if α is close to one there is a higher probability that the traffic rate is less than EB.
- (2) Decreasing EB, on the other hand, will increase the delay exponentially for a given α . The formula bandwidth = traffic/time can no longer be used to calculate a network bandwidth to carry fractal traffic.

4.0 CONCLUSION

Network efficiency is determined by traffic characteristics, e.g., the peak-to-average rate ratio. Dramatic traffic burstiness affects network utilization. Without traffic control, fractal traffic is bursty in multiple timescales. Simple multiplexing does not smooth out such burstiness rather than causes traffic congestions and packet losses. Fractal traffic needs to be measured and controlled to achieve higher network efficiency. A fractal traffic flow has an efficient bandwidth (EB). Above this bandwidth, there is limited improvement of performance. Large packet-level bursts may overflow the buffer at the bottleneck link and cause packet losses. The observed fractal nature of the internet will be modeled using the concept of L-System in future [11].

ACKNOWLEDGEMENT

We are extremely thankful to the editorial board of this journal for inclusion of this research work.

References

- [1] W. E. Leland, M. S. Taqqu, W. Willinger, D. V. Wilson, "On the Self-Similar Nature of Ethernet Traffic," *IEEE/ACM Trans. Networking* Vol. 2, pp.1-15. 1994.
- [2] V.Paxson and S.Floyd, "Wide-area traffic: The failure of Poisson modeling," *IEEE/ACM Trans. Networking*, vol.3, pp.226-244, 1995
- [3] Patrice Abry, Richard Baraniuk, Patrick Flandrin, Rudolf Riedi, Darryl Veitch, "The Multiscale Nature of Network Traffic: Discovery, Analysis, and Modelling," *IEEE Signal Processing Magazine*, May 2002

- [4] M.E.Crovella and A.Bestavros, “Self-similarity in world wide Web traffic – evidence and possible causes,” *Proc. ACM Sigmetrics’96*, pp.160-169, 1996.
- [5] K.Park and W.Willinger, Eds., *Self-Similar Network Traffic and Performance Evaluation*. Wiley Interscience, 2000.
- [6] K.Park, G.Kim, and M.Crovella, “On the relationship between file sizes, transport protocols, and self-similar network traffic,” *Proc. IEEE International Conference on Network Protocols*, pp.171-180, 1996.
- [7] A.Feldmann, A.C.Gilbert, and W.Willinger, “Data networks as cascades: Investigating the multifractal nature of Internet WAN Traffic,” *Proc. ACM/Sigcomm’98*, pp.25-38, 1998.
- [8] A.Feldmann, A.C.Gilbert, P.Huang, and W.Willinger, “Dynamics of IP Traffic: A study of the role of variability and the impact of control,” *Proc. ACM/Sigcomm’99*, pp.301-313, 1999.
- [9] A. Feldmann, A.C.Gilbert, W.Willinger and T.G.Kurtz, “The changing nature of network traffic: Scaling phenomena.” *Computer Communication Review* 28, No, April 1998
- [10] Z.-L. Zhang, V.Ribeiro, S.Moon, and C.Diot, “Small-time scaling behaviors of internet backbone traffic: an empirical study,” in *Proc. IEEE Infocom*, San Francisco, CA, Apr. 2003.
- [11] Mishra, S.N. et al: *L-System Fractals*, Book, Mathematics in Science and Engineering, Vol-209,Elsevier, Netherlands,2007.