Emergent Phenomena in the Internet

Adithya Kuchibhotla
Physics 569 Term Essay

December 19, 2012

Abstract

In this essay, a brief structure of the Internet is first given and then two emergent phenomena that are exhibited by the Internet, viz., Self-similar traffic and Scale-free topology are studied. Suitable reasoning behind their occurrence and some mathematical modelling is also given. The Barabasi-Albert model that tries to explain the Scale-free Topology of Internet is explained. Finally, criticisms of some of these models are given along with the reasons for necessity to study the Internet phenomena further in the future.

1 Introduction

The Internet, defined as a global collection of computer networks that communicate through Internet Protocol (IP), can be described as a "prime example of a large-scale, highly engineered, yet highly complex system" [1]. It is a complex system in the sense of a complicated system exhibiting simple behaviour [2]. The Internet is by far the largest man-made structure and continues to evolve in a self-organised way. It also offers a ground to study various correlation phenomena that occur in complex systems and to test models that describe them. Hence, the need to study the dynamics of the Internet is increasing day-by-day. The Internet is heterogeneous and exhibits emergent phenomena that depend on its large-scale organization. Here, we try to describe two such phenomena the Internet exhibits, viz., Self-Similar traffic and Scale-free topology.

2 Structure and Technology of the Internet

The Internet at the largest scale can be divided into Automated Systems(ASs) that communicate with each other via the IP. Each AS is a collection of routers and links. The Internet has a hierarchic structure of routing with the Internet Service Providers (ISPs) considered to be at the bottom. The IP is the transport layer that unifies thousands of networks and is responsible for the transport, routing and delivery of each single packet of data. It is a packet oriented service in which the total bandwidth is available for each of the packets to compete with the others.

The layer above IP is the Transmission Control Protocol (TCP) which controls congestion through its additive-increase/multiplicative decrease congestion control mechanism. The layer above TCP is the top layer which is the application layer that serves the user directly. It contains a range of protocols such as FTP, HTTP, SMTP, etc.

3 Self-Similar Internet Traffic

The traffic rates in the Internet, when measured, exhibit self-similar or fractal-like behaviour, i.e., a segment of the traffic rate process measured at some time scale looks or behaves like a scaled version of the traffic rate process measured over a different time scale [3][4][5]. The measurements discovered autocorrelations that decay according to a power law and not as an exponential law as traditionally believed. If the traffic were to follow a Poisson process, "it would have a characteristic burst length which would tend to be smoothed by averaging over a long enough time scale" [3].

3.1 Mathematical Modelling

Since a self-similar process has observable bursts at a wide range of time scales, it can exhibit long-range dependence. Consider a second-order stationary and zero mean stochastic process $X = (X_k : k \geq 1)$ with autocorrelation function $(r(k), k \geq 0)$ and define the family of aggregated processes $(X(m) : m \geq 1)$, where, for $m = 1, 2, \cdots, X^{(m)} = (X^{(m)}(i) : i \geq 1)$ is given by

$$X^{(m)}(i) = \frac{(X_{(i-1)m+1} + \cdots + X_{im})}{m}. \tag{1}$$

X is called asymptotically second-order self-similar (with self-similarity or Hurst parameter $0 < H < 1$), if [3],

1. $\lim_{m\to\infty} \mathrm{Var}(m^{1-H}X^{(m)}) = \sigma^2$, where $0 < \sigma^2 < \infty$ is a finite positive constant, and
2. $\lim_{m\to\infty} r^{(m)}(k) = ((k+1)^{2H} - 2k^{2H} + (k-1)^{2H})/2$, where $r^{(m)} = (r^{(m)}(k), k \geq 0)$ denotes the autocorrelation function of the aggregated process $X^{(m)}$.

Now, X is said to exhibit Long Range Dependence if for $1/2 < H < 1$,

$$r(k) \sim k^{2H-2}, \text{ as } k \to \infty \tag{2}$$

Different methods can be employed to obtain the value of H, [6] such as,

1. The variance-time plot method, in which the variance of $X^{(m)}$ is plotted against m on a loglog plot; a straight line with slope greater than is indicative of self-similarity, and H is given by $H = 1 - \beta/2$.
2. The R/S plot method, in which,the rescaled range or R/S is plotted against the number of points n a loglog plot, whose slope gives an estimate of H.
3. The periodogram method, in which the the slope of the power spectrum near frequency zero in a log-log plot gives the value of H by the relation Slope= $1 - 2H$.

3.2 Experiment-Measurement of Internet Traffic

The following plot based on work in ref.[7] shows the measurement of self-similarity in the Internet traffic.

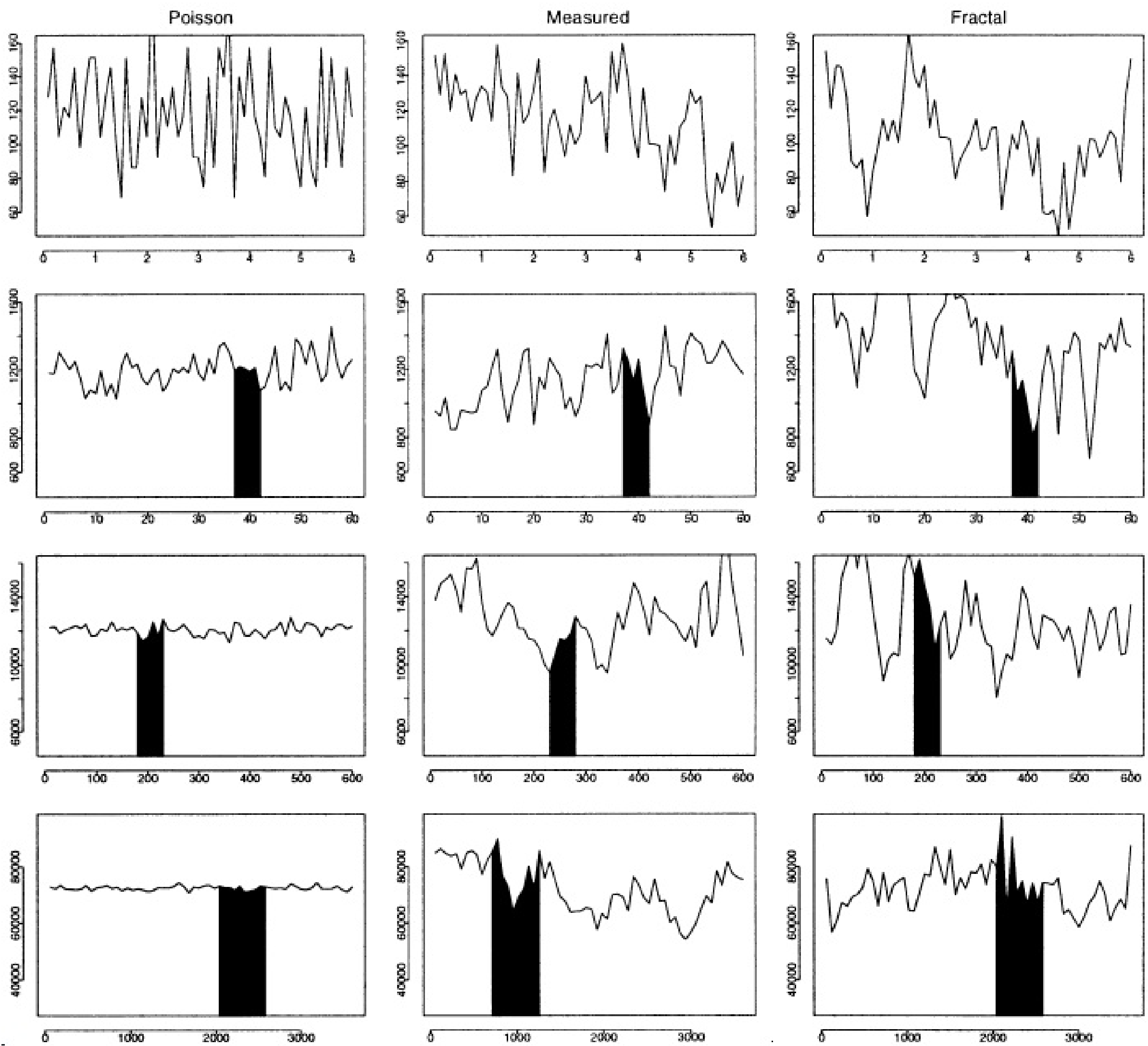

Figure 1. Internet traffic viewed over four orders of magnitude.

Each plot shows the number of packets recorded during a randomly selected subset of a trace of Internet traffic. The central column shows the measurement of the trace. The top plot of the measurement shows a randomly selected subset of the trace on a time scale of 100 msec for a total of 6 sec. The second plot shows a time scale that is a factor of 10 larger so that each observation represents the number of packets per 1 sec, spanning 60 sec in total. The black-shaded region indicates from where the plot in the row above was chosen. This process is repeated to two more factors of scaling. It can be clearly seen that the measured plots resemble plots of fractal-like model (the plots on the right column) while the plots of Poisson model (the plots on the left column) tend to smooth out quickly at larger time scales. The Internet traffic, remains invariant and shows bursty behaviour at all time-scales.

Another measurement done using the methods to find H described above yields the following plots [3]:

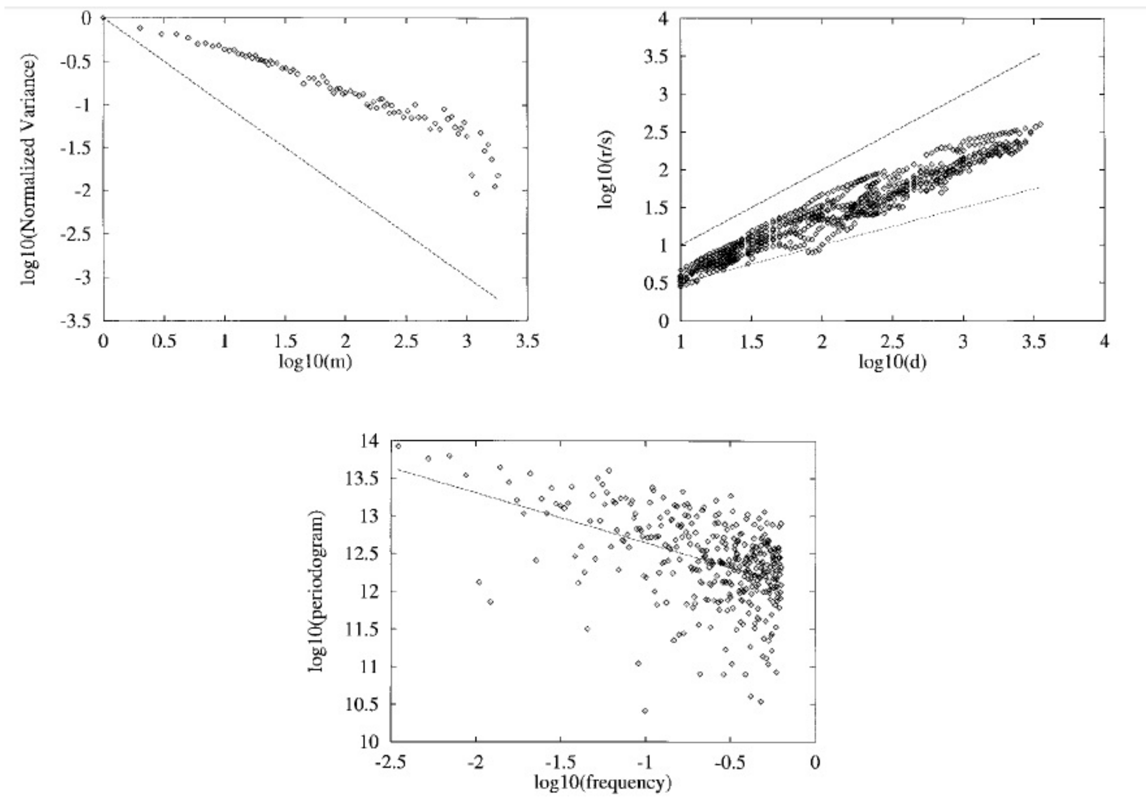

Figure 2. Graphical analysis of Internet traffic data

The figure shows plots for the three graphical methods: variance-time plot (upper left), R/S plot (upper right), and periodogram plot (lower center). The variance-time plot is linear with a slope of 0.48 giving an estimate for H of 0.76. The R/S plot shows an asymptotic slope of 0.75 which gives an estimate of H of 0.75. The periodogram plot shows a slope of 0.66, giving an estimate of H as 0.83. Hence we get a value of H that is $1/2 < H < 1$ which shows the Long Range Dependence.

4 Scale-free topology

The protocol in the Internet's structure that is responsible for determining the path along which the data packets are to be sent is the Border Gateway Protocol (BGP) which connects different ASs. The Internet can be decomposed into connected subnetworks that are under separate administrative authorities, as can be shown in Figure 3 [4].

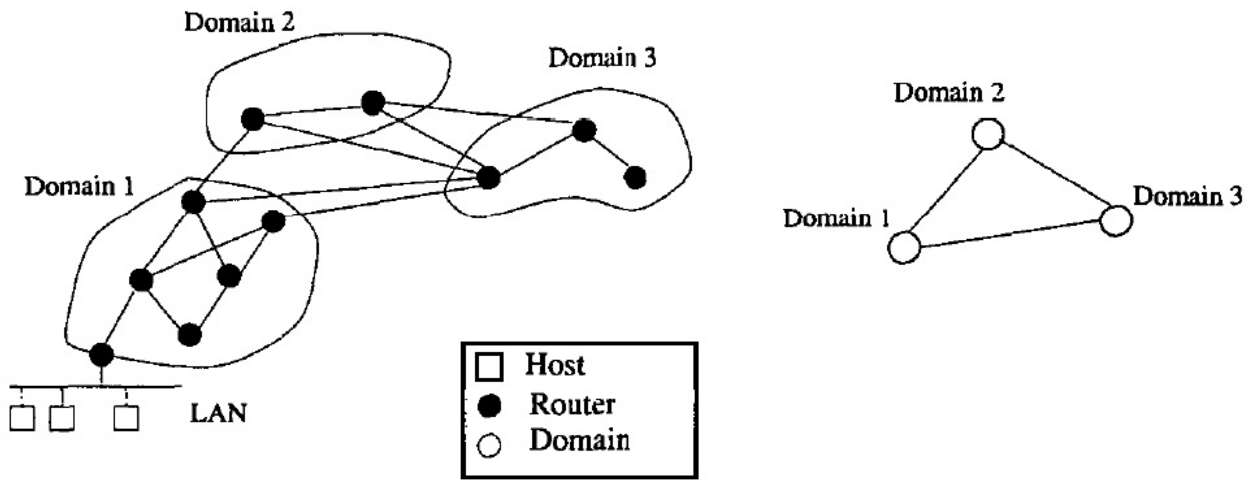

Figure 3. The structure of Internet

The figure shows a graphical structure of the Internet at the router level and at the inter-domain level. The number of different other vertices a vertex is connected to is it's degree. It is discovered that there is a power-law behaviour of the vertex distribution of snapshots of the AS-connectivity graph, which means, at any time, there is a considerable probability of finding a vertex in the graph with degree greater than, say, one or two. This implies the non-existence of 'preferred' vertices and hence the absence of a characteristic scale.

By the power law behaviour of the vertex degree distribution of the AS graphs, it is meant that if f_d is the number of vertices with degree d, then $f_d \propto d^{-\alpha}$ for some α. Equivalently, in terms of a complimentary distribution function, this can be expressed as $1 - F_d = 1 - \sum_{i=1}^{d} f_i \propto d^{-(\alpha-1)}$. Figure 4 shows this behaviour.

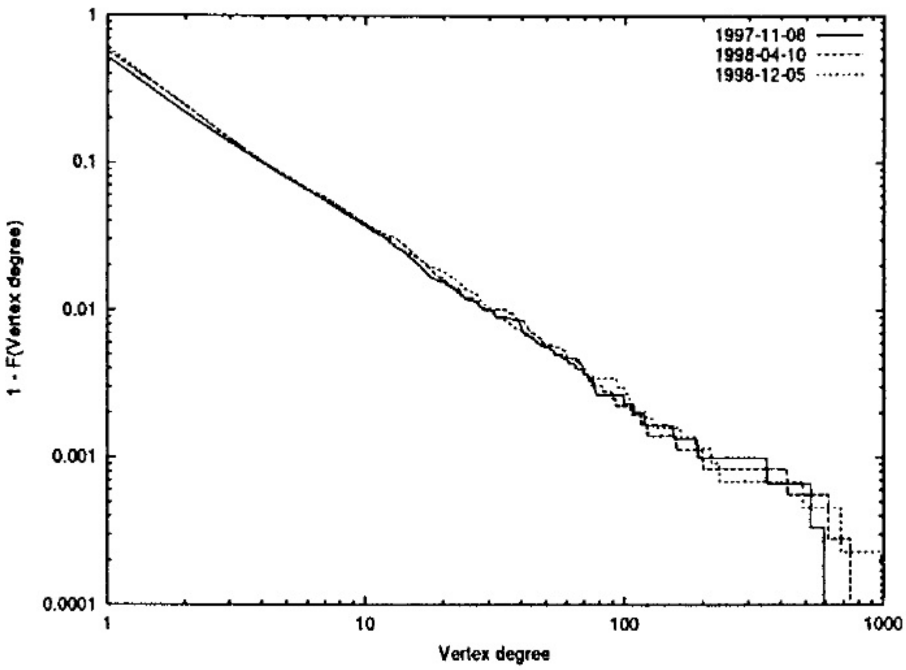

Figure 4. Log-Log plot of 1-F_d vs. d for three different AS graphs

For each graph, α is found to be ~ 2.1, which means the vertex degrees observed in an AS graph is highly variable which implies the absence of any preferred degrees in the graph around which the plots are concentrated. Hence, these power-law graphs are also called scale-free graphs.

4.1 The Barabasi-Albert Model

Barabasi, Albert and colleagues [8][9] tried to explain the scale-free graphs by constructing models based on three main mechanisms:

1. Incremental growth, which follows from the observation that most networks develop over time by adding new nodes and new links to the existing graph structure.
2. Preferential connectivity, which takes into account the higher probability for a new or existing vertex to connect to a vertex that has high degree than there is to connect to a vertex of low degree.
3. Rewiring, which accounts for randomness or noise in the formation of networks by replacing certain links in the graph.

Authors of [8][9] were able to show that graphs constructed using these mechanisms has a vertex-degree distribution that follows power law.

5 Criticism

The self-similar behaviour of the Internet traffic, as described before, is observed both in local-area networks [10] and in wide-area networks [3][11]. Traffic modellers and Internet researchers put forward models to explain the self-similarity but work of [1] has shown that these models are “evocative” and not “explanatory” in the sense that the models do not allow for tests on the assumptions the models are based up on.

Also, the authors of [1] have worked on the mechanisms that the Barabasi-Albert model is based on and found that new ASs connect to an existing graph in a different way that the third mechanism in the Barabasi-Albert model. Their work on analyzing how new ASs connect using Internet traffic data is shown in Figure 5.

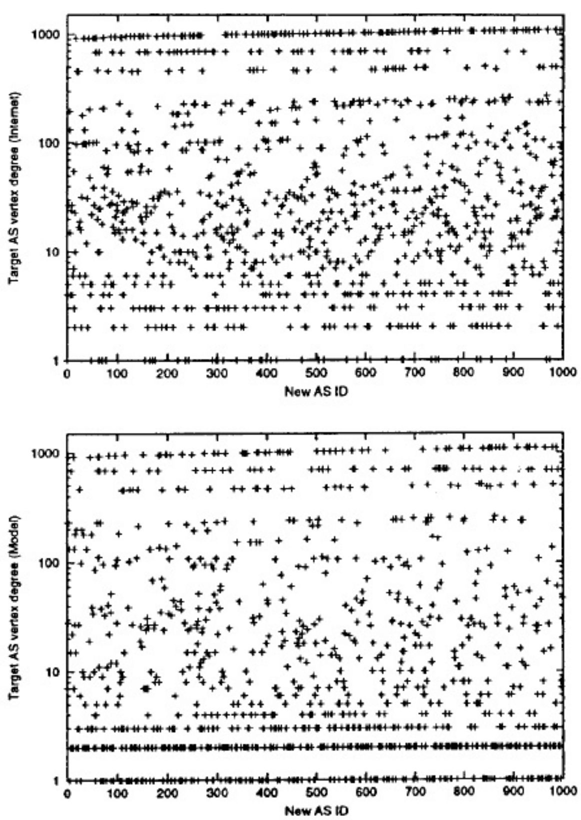

Figure 5. The target-vertex degree for each new AS created.

Form the figure, it can be seen that the new ASs prefer higher degree vertices more than what the BA model predicts. Hence, the authors claim that the BA model is not explanatory and that new explanatory models are required to explain the measurements.

6 Conclusions

We have seen how Internet exhibits the emergent phenomena, Self-Similar behaviour of Internet traffic and Scale-free graphs of interconnecting ASs' graphs that define the Internet topology. We have also seen the measurements made that establish the fact that Internet exhibits these phenomena. Various explanations and their criticisms are also studied. Hence, we can conclude that the Internet is growing to be a very interesting phenomenon itself with lot of opportunities at finding interesting phenomena within itself. Also, there is much work that needs to be done that will help us in understanding the behaviour of the dynamics of the Internet.

References

[1] Willinger, W., Govindan, R., Jamin S., Paxson V., and Shenker S. *Scaling phenomena in the Internet: Critically examining criticality*,

[2] Pastor-Satorras, R., Vespignani A. *Evolution and Structure of the Internet: A Statistical Physics Approach*, 2004, Cambridge University Press.

[3] 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, December 1997.

[4] Faloutsos M., Faloutsos P. Faloutsos C., *On Power-Law Relationships of the Internet Topology*, Comput. Commun. Rev. 29, 251-262.

[5] Waxman, B. M. (1988) IEEE J. Select. Areas Comm. 6, 1617-1622.

[6] Beran J., *Statistics for Long-Memory Processes (Monographs on Statistics and Applied Probability)*, London: Chapman and Hall, 1994.

[7] Willinger W., Paxson, V., *Where Mathematics Meets the Internet*, 1998, Notices of the AMS45, Number 8.

[8] Albert, R., Barabasi A-L., (2000) Phys. Rev. Lett. 85, 5234-5237.

[9] Albert, R., Jeong, H., Barabasi A-L., (2000) Nature 406, 378-382.

[10] Leland, W. E., Taqqu, M. S., Willinger, W., Wilson, D. V. (1994) IEEE/ACM Trans. Network. 2, 1-15.

[11] Paxson, V., Floyd, S. (1995) IEEE/ACM Trans. Network. 3, 226-244.