Traffic Self-Similarity

Adrian Popescu
Department of Telecommunications and Signal Processing
Blekinge Institute of Technology
371 79 Karlskrona, Sweden

The unifying concept underlying fractals, chaos and power laws is self-similarity. Self-similarity, or invariance against changes in scale or size, is an attribute of many laws of nature and innumerable phenomena in the world around us. Self-similarity is, in fact, one of the decisive symmetries that shapes our universe and our efforts to comprehend it.

Manfred Schroeder, 1991

1 Introduction

Statistical analysis of high-resolution traffic measurements from a wide range of working packet networks (e.g., Ethernet LANs, WANs, CCSN/SS7, ISDN and VBR video over ATM) have convincingly shown the presence of fractal or self-similar properties in both local area and wide area traffic traces [2, 3, 5, 15, 19]. That means that similar statistical patterns may occur over different time scales that can vary by many orders of magnitude (i.e., ranging from milliseconds to minutes and even hours). This discovery calls to question some of the basic assumptions made by earlier research done in control, engineering, and operations of broadband integrated systems. The fact that network traffic is inherently fractal or long-range dependent (LRD) and many studies assume traffic to be short-range dependent (SRD) leads one to question the extent to which the results of these studies are applicable in practice. At the time being, there is mounting evidence that LRD is of fundamental importance for a number of teletraffic engineering problems, such as traffic measurements [24], queueing behaviour and buffer sizing [6, 16], admission control [4] and congestion control [21]. Unlike traditional packet traffic models, which give rise to exponential tail behavior in queue size distributions and typically result in overly optimistic performance predictions and inadequate resource allocations, fractal traffic models predict hyperbolic or Weibull (stretched exponential) queue size distributions and could therefore affect control and management of present and future packet networks within an integrated framework of end-to-end Quality of Service (QoS) provision. This also indicates that increasing link capacity could be more effective in improving performance than increasing buffer size.

Although similar processes have been observed and analysed in a number of other areas like, for instance, hydrology, biophysics, financial economics [23], the work on fractal dynamics of network traffic behaviour is relatively new. While most of the work done in science and engineering has almost exclusively focused on statistical and practical features of fractal models (e.g., data analysis, mathematical modeling), the engineering impacts on performance and analysis have not yet received an adequate interest. This is mainly because of the difficulties related to analysis and the ability to use these models in control. Self-similarity implies that a specific correlation structure is retained over a wide range of time scales, albeit it may come in different forms. In order to be able to do control on such processes, one needs first to explain and to validate on physical grounds the causal mechanisms that could be responsible for generating self-similarity in a realistic network environment. Furthermore, understanding the impacts of such processes is at least as important as the understanding of their physical origins.

2 Traditional and non-traditional truths

Self-similarity is a concept related to fractals and chaos theory. The recent history of mathematics and natural sciences, to capture features of natural phenomena, testifies to the revision of several important traditional truths [23]. Features of natural, complex phenomena include, among others, out-of-equilibrium dynamics, unpredictability, non-linearity, feedback loops, phase transitions, attractors, adaptation, and inductive behavious.

For instance, one of the first traditional truths (20th century) takes the form of ”quantitativeness”, i.e., most of physical theories are and should be quantitative. However, later events, like for instance the upcoming of catastrophe theory (which deals with forms of things and not their magnitudes), has turned this truth into a non-traditional one. Qualitative aspects have been found to be as important, and sometimes even more important, than the quantitative ones. Accordingly, many natural phenomena can no longer be represented by means of analytic functions (traditional truth) but they are usually singular in character and therefore difficulties in representation by analytic functions (fractals). Their evolution, although derivable from dynamical equations, can no longer be predicted from motion equations, as they are not necessarily predictable (chaos).

Furthermore, another important traditional truth, according to which physical systems can be characterized by so-called fundamental scales (e.g., in time and in length), has turned into a non-traditional one, namely the scaling principle. That means, natural phenomena (whose central moments diverge such as inverse power law) do not necessarily have fundamental scales to describe their behavior but, on the contrary, they may be described by scaling relations. For instance, the length of a fractal curve depends on how it is measured, i.e., the scale of the measuring instrument.

The traditional principle of superposition has been also shown to be violated by most natural phenomena, because of diverse reasons like non-linearities, perturbations, etc. The consequence is that understanding a process as a whole cannot be accomplished only by the knowledge of the decomposed elements of its behavior. The evolution of natural phenomena may become even more complex than that according to the superposition principle.

3 Fractals and self-similarity

A self-similar phenomenon represents a process displaying structural similarities across a wide range of scales of a specific dimension. In other words, the reference structure is repeating itself over a wide range of scales of diverse dimensions (geometrical, or statistical, or temporal), and the statistics of the process do not change with the change. However, these properties do not hold indefinitely for real phenomena and, at some point, this structure breaks down.

Self-similarity can therefore be associated with ”fractals”, which are objects with unchanged appearances over different scales. The concept of fractals includes, besides the geometrical meaning, statistics as well as dynamics. That means, there are fractal processes of diverse dimensions, e.g., geometrical, statistical and dynamical. Examples of geometrical fractal processes are the Cantor set, Sierpinski triangle, Koch curve, etc [20].

In the case of statistical fractals it is the probability density that repeats itself on every scale, like for instance as found in economics (Pareto's law), in linguistics (Zipf's law) and in sociology (Lotka's law) [23]. On the other hand, a dynamical fractal is generated by a low-dimensional dynamical system with chaotic solutions, as it happens in biology (Willis law) and in medicine (cardiac spectrum, mammalian lung).

4 Chaos and randomness

The concepts of ”chaos” and ”order” have been long time viewed as antagonistic, with the consequence of development of specific investigation methods. Diverse ”natural” laws (e.g., Newton's law, Kepler's law) represent good examples of ”ordered” nature, whereas chaos belongs to a distinct part of the nature where simple laws are no more valid. In other words, chaos does not necessarily mean a higher degree of complexity, but a condition when nature fails to obey laws with different degrees of complexities [20]. Furthermore, things may get even more challenging based on the observation that natural systems seem to have no difficulties in switching from one state to another.

A chaotic system represents a system with an extreme sensitive dependence on initial conditions. Very small changes in (diverse) initial conditions (e.g., specific parameters, background noise, innacuracy of equipment) may drastically change the long-term behaviour of the system. For instance, with an initial parameter of, say, 5.0, the final result of a chaotic system may by entirely different from that of the same system with an initial parameter value of, say, 5.0000001.

Chaos, as a meteorological phenomenon, was initially discovered by a meteorologist, Edward Lorenz, in 1960 [20]. Based on numerous studies, he stated that it is impossible to forecast the weather accurately. However, his studies led finally to other issues of what eventually got to be known as the chaos theory (the Lorenz attractor, interplay between randomness and deterministic fractals, etc). Other examples of chaotic systems have been subsequently found in ecology, prediction of biological populations (the well-known bifurcation diagram), human heart, computer sciences, fractal images (Mandelbrot set, Sierpinski triangle, Koch curve), etc [20]. Some of the most fundamental characteristics of these systems are the fact that randomness creates finally deterministic shapes (trajectories) as well as the existence of the so-called invariants, i.e., different constants that are universal valid (e.g., the Feigenbaum constant, diverse invariants that characterize the web server workloads, etc). The consequence therefore is that nothing in nature seems to be random, and a process that we experience as being random is perhaps so only because of the incompletness of our knowledge.

5 Traffic measurements

In 1993, a group of researchers from Bellcore and Boston University delivered a paper at SIGCOMM'93 conference that is entitled "On the Self-Similar Nature of Ethernet Traffic" [14]. An extended version of this paper was also published the following year in the *IEEE/ACM Transactions on Networking* [15]. These papers are perhaps some of the most important papers of the decade produced in the area of computer networks. They have also triggered the upcoming of many other studies and reports on diverse traffic measurements in a wide variety of networking situations. The general conclusion has been that statistical analysis of high-resolution traffic measurements from a wide range of working packet networks (e.g., Ethernet LANs, WANs, CCSN/SS7, ISDN and VBR video over ATM) have convincingly shown the presence of self-similar properties in both local area and wide area traffic traces [2, 3, 5, 19].

A short description of the models found to best describe traffic traces suitable for a number of specific applications is as follows.

5.1 Ethernet traffic

Using a massive amount data of Ethernet traffic measurements collected between 1989 and 1992 from various Ethernet Local Area Networks (LANs) at Bellcore, the authors of the above-mentioned papers (W. Leland, M. Taqqu, W. Willinger and D. Wilson) have demonstrated that the (aggregate) Ethernet traffic is self-similar. Four different sets of traffic measurements are reported, each containing 20 to 40 hours of Ethernet traffic, and with a total of over 100 millions packets. Furthermore, detailed analysis (aggregation, autocorrelation, R/S analysis, variance-time plot, periodograms, Whittle's estimator, etc) and very rigorous statistical analysis (confidence intervals, sophisticated statistical tests, etc) have been conducted to understand and to explain the observed, realistic traffic patterns as well as to model these patterns.

Some of the main results reported are:

- Poisson (or Markovian) models do not capture reality
- Aggregate Ethernet LAN traffic is self-similar
- Burstiness is observed accross many time scales
- Hurst parameter: $0.7 < H < 0.9$
- The parameter H is larger when network utilization is higher
- Traffic do not aggregate well
- The law of large numbers may not hold

Many other papers have been ulteriorly published that have confirmed the self-similar nature of data traffic in Ethernet and other networking environments. A physical explanation of the self-similarity observed in the Ethernet traffic is also provided in [22, 25].

5.2 Wide Area Networks traffic

In 1994, two researchers from the Lawrence Berkeley Laboratory (V. Paxson), and from the University of California, Berkeley (S. Floyd) published a paper at the SIGCOMM'94 conference that is entitled "Wide-Area Traffic: The Failure of Poisson Modeling" [18]. An extended version of this paper was also published the following year in the *IEEE/ACM Transactions on Networking* [19]. In these papers, the authors report the results of measurement and analysis studies done on WAN traffic (TCP traffic), for applications like FTP and TELNET. Millions of TCP packets and connections have been collected from different sites with traces ranging from 1 hour to 30 days. Furthermore, detailed measurement studies have been done (lengthy Internet packet traces, high resolution timer, lots of storage space, etc) as well as detailed analysis (per application, connection level, connection interarrivals, etc) and very rigorous (confidence intervals, sophisticated statistical tests, etc) statistical analysis.

Some of the main results reported are:

- Poisson seriously underestimates the burstiness of TCP traffic over a wide range of time scales
- Self-similarity is present in aggregate WAN traffic
- TELNET connection arrivals appear to be Poisson, but packet arrivals are not
- FTP connections do not appear to be Poisson. FTP session arrivals correspond well to Poisson, but data connection arrivals have a much burstier arrival rate. Furthermore, the number of bytes per burst seems to have a heavy-tailed distribution

Similar to the Ethernet studies, the papers on WAN traffic have been followed by many other measurement and analysis studies that have confirmed the self-similar nature of WAN traffic.

5.3 Variable Bit Rate video traffic

Several studies have been published that have shown the self-similar nature of the digitized (VBR) video traffic as transmitted over ATM and Internet. The most important publications are those produced by M. Garett and W. Willinger at SIGCOMM'94 ("Analysis, Modeling, and Generation of Self-Similar VBR Video Traffic") [10] and by J. Beran, R. Sherman, M. Taqqu and W. Willinger in 1995 in the *IEEE Transactions on Communications* (”Long-Range Dependence in Variable-bit-rate Video Traffic”) [2].

Some of the main results reported are:

- Video transmission exhibits self-similarity
- Frame length conforms to Pareto distribution, at least in the tail of the distribution
- LRD seems to be an inherent feature of VBR video traffic, no matter the codecs used and the scene recorded

5.4 World Wide Web traffic

A group of researchers from the Boston University (M. Crovella, A. Bestavros and others) published a set of papers in 1995 and 1996 where they report measurement studies done of WWW traffic, and where the main result was that this traffic is self-similar [3]. Over a half million Web request traces from about 40 workstations at the Boston University have been collected and analyzed. The study was divided into three classes of measurements, namely on total WWW traffic, individual user (source) behavior and transmission times. The authors did also additional tests to try to explain the self-similar behavior.

Some of the main results reported are:

- Traffic patterns generated by browsers have a self-similar nature
- Every Web browser is modeled as an ON-OFF source model and data fits well the Pareto distribution
- Files available via the Web over the Internet seem to have a heavy-tailed size distribution (bimodal distribution)
- Attempts have been done to explain the self-similar behavior of the traffic

For instance, the authors have observed that the degree to which file sizes are heavy-tailed can directly determine the degree of traffic self-similarity, and this is not significantly affected by diverse parameters like variations in bottleneck bandwidth and/or in buffer capacity, cross-traffic and the distribution of interarrival times.

Furthermore, an important work is also reported by researchers from the University of Saskatchewan, who did measurement and analysis studies on Web server workload patterns [1]. They have finally identified a number of ten workload invariants, like success rate, mean transfer size, concentration of references, size distribution etc.

Finally, other important results are reported in [11]. The authors show, among others, that the distribution of HTTP document sizes can be modeled by a mixure of Lognormal and Pareto distributions as well as that diverse classes of Web servers (e.g., academic, entertainment, commercial) seem to show structural similarities in their distributional properties.

5.5 Signaling System Number 7 traffic

Measurement and analysis studies on SS7 traffic have been done at the Dublin Institute for Advanced Studies [5]. The authors have collected, and analyzed, more than 170 million signaling messages from a variety of different SS7-controlled networks.

The main results are:

- Poisson models do not capture the SS7 traffic patterns
- SS7 traffic is self-similar

Finally, it is also mentioned that the list of applications showing self-similar properties is not limited to the above-mentioned applications. Still, there are other applications that show a self-similar behavior as well (e.g., SMTP, deterministic data transfer) [11].

6 Origins of Self-Similarity in Teletraffic

Recent studies have shown that self-similarity in an *idealized* environment (with unbounded resources and independent traffic sources) can derive from the aggregation of many individual, albeit higly variable, ON-OFF sources (i.e., ON- and OFF-periods with heavy-tailed distributions with infinite variance, e.g., Pareto distributions) [2, 3, 15, 18]. In other words, the superposition of many ON-OFF sources exhibiting the infinite variance syndrome (the so-called ”Noah Effect”) results in self-similar aggregate network traffic approaching fractional Brownian motion (the so-called ”Joseph Effect”). Furthermore, investigation of diverse traffic sources have also revealed that the highly variable ON/OFF behavior is an intrisic property of the client-server interaction [1, 3, 11].

These results, although valuable and elegant, suffer however because of assumptions that are not realistic in terms of network environment. Contrary to the above-mentioned assumptions, *realistic* client/server network environments implies resource boundedness. That means that non-linear processes could be also induced because of competition for bounded resources, with the consequence of coupling among traffic sources. Furthermore, diverse (feedback-based) control mechanisms, e.g., OS scheduling algorithm, TCP, Ethernet, may induce additional non-linearity in the case of congestion as well.

The complexity of understanding the underlying physics that could give rise to self-similarity in network traffic is mainly given by the fact that this is not induced by one physical phenomena, but more. Different correlations existent in a self-similar network traffic, which are acting at different temporal scales, may arise because of different physical phenomena that manifest themselves with characteristics relevant to the specific temporal scale, e.g., information organization and retrieval (applications, disk and memory scheduling), user ”think time” and preference in file transfers (session/activity), effects of caching, TCP, Ethernet, and diverse ATM control mechanisms (admission control, congestion control, etc.). Furthermore, difficulties in understanding the interactions between diverse ”individual” correlations created in network traffic by these physical phenomena further complicates the issue.

For instance, some of the most significant physical phenomena that may give rise to effects of different kinds in network traffic are as follows:

- User behavior
- Data generation, organization and retrieval
- Traffic aggregation
- Network controls
- Feedback-based control mechanisms
- Network evolution

6.1 User behavior

One of the most important factors that affects the traffic behavior (at a session/call level and also during the session) is due to (human) user behavior. It was found, for instance, that the distribution of user requests (think time) and preferences for documents on Internet (WWW) show an extreme degree of fluctuations over a wide range of time scales [3, 11]. Furthermore, different flow control mechanisms existent in diverse traffic sources (e.g., VBR video MPEG-coded sources, ABR, TCP), which regulate the output traffic rate depending upon the states of the network, may also contribute to an increased burstiness in the network traffic [12]. It is also relevant to mention that the distributions of many phenomena of potential social and biological implications were shown to have inverse power law tails [23]. Examples of such systems are found in sociology (Lotka’s law), linguistics (Zipf’s law), economics (Pareto’s law), medicine (cardiac spectrum) and biology (Willis’ law).

6.2 Data generation, organization and retrieval

An important aspect concerning the origins of traffic self-similarity is related to the way data is processed. Measurement studies have shown that self-similarity in network traffic as well as in application traffic is a two-dimensional property that refers to the interarrival time distributions and also to the size distributions of information entities. These results suggest that, at least for the application traffic, self-similarity may not be a machine induced artifact. Different applications/sources may supply traffic of statistically distinctive characteristics, but key statistical behavior (e.g., correlation structure) may be invariant from one machine/network to another.

6.3 Traffic aggregation

One of the most serious obstacles that prevents from keeping the traffic generated by individual sources isolated from other traffic (at least for as far into the network as possible) is due to the actual form of traffic aggregation (i.e., statistical multiplexing), used in the packet- and cell-based networks, with the consequence of coupling among traffic sources. As mentioned above, superposition of many ON-OFF sources exhibiting the infinite variance syndrome results in self-similar aggregate network traffic approaching fractional Brownian motion. The LRD characteristics in traffic are also shown to be highly robust with respect to a variety of network operations, such as traffic splitting, merging, queueing, policing and shaping [6]. Self-similarity seems to be preserved under superposition of both homogeneous and heterogeneous, i.e., independent, traffic sources, and this holds over a wide range of conditions like for instance variations in the bandwidth bottleneck and buffer capacity, and mixing with cross-traffic possessing dissimilar (correlation) characteristics [12, 17]. Once a particular traffic source generates LRD traffic, the aggregate network traffic seems to become LRD, irrespective of the characteristics (SRD or LRD) of other traffics in the mix. The coupling process is a very complex one, an entire range of properties might get altered, i.e., not only the Hurst parameter, but also the mean and the variance might get significantly changed [6, 7].

6.4 Network controls

One of the main reasons behind traffic complexity seems however to be the resource boundedness that is typical in a realistic network environment, e.g., bounded communication and switching resources/bandwidth, limited buffer space, finite (CPU) processing capabilities. This is to be judged in the light of heavy-tailed asymptotics of LRD traffic processes existent in such environments, which generally claim more resources in the case of traffic control. Non-linear processes could be induced in the case of control mechanisms where diverse conflicts are not solved properly simply because of bounded resources. This is a general problem, which could manifest in many forms, from simple queueing models to very sophisticated control mechanisms. Furthermore, the problem of control of LRD traffic may even get exacerbated in the case of competition of more users for bounded resources, in the case of which the problem of correct resource dimensioning becomes of paramount importance. Finally, this issue is further complicated because of its multi-dimensional character, e.g., hardware contention (CPU, memory), bandwidth contention, and software contention (OS scheduling strategies, process priorities).

6.5 Feedback-based control mechanisms

A further complication is because many control mechanisms are of feedback-based type, e.g., flow and congestion control mechanisms (TCP, rate-based control mechanisms, etc). That means additional non-linearity could be induced in the case of congestion, with possibilities for a wide range of dynamical system behaviour, such as multistability and chaos [13]. Multi-fractal scaling behavior of transport protocols are especially relevant [9]. Furthermore, it is also important to notice that there may exist very complex interactions between workload fluctuations and diverse (network) control mechanisms. These interactions not only include the potential for controls to temper the fractal nature of traffic (e.g., traffic shaping, UDP) but, in some cases, may also create or even enhance it (e.g., OS scheduling contention, TCP, Ethernet) [6, 17]. Two sets of critical issues are identified in this case, namely, one on the impact of realistic traffic on the efficacy of specific traffic control mechanisms, and the other dealing with the extend to which controls may modify traffic characteristics.

6.6 Network evolution

Finally, an important reason for the increase of burstiness in the network traffic is because of the network evolution, with the continuous appearance of new services and applications. Especially the upcome of WWW-based services has dramatically resulted in more sophisticated traffic patterns [9].

As a quantitative estimation, link layer characteristics (e.g., Ethernet) seem to dominate the network traffic profile at millisecond time scales [15]. In contrast, the effects of human behaviour as well as data generation and retrieval seem to dominate at tens of seconds and beyond (minutes and even hours) [3]. In between, the effects of diverse flow control mechanisms, such as TCP, would likely dominate (at time scales of seconds). Queueing characteristics might dominate at tens and hundreds of milliseconds as well [6].

7 Performance Impacts

Understanding and validating self-similarity on physical grounds in a realistic network environment is, although important, not enough when developing efficient and integrated network frameworks within which requested end-to-end QoS guarantees are fully supported. An equal importance must be placed on understanding the impacts of self-similarity as well. These impacts can be broadly partitioned into fractal queueing (i.e., queueing models driven by fractal traffic processes) and impacts on controllability [6, 7].

Packet network traffic in LAN and WAN environments has been shown to be LRD. Packets enter the network in clusters. Because of the LRD phenomenon, the bursts do not smooth out by simply aggregating over longer timescales. Such traffic is an inherent property of the sources that alternate between ON and OFF states where at least one of the states has heavy-tailed distribution for the time spent in that state. The LRD emerges as a manifestation of the high variability in the ON and OFF periods.

Almost all applications on the Internet follow the client-server paradigm. The client makes a REQUEST for a service at the server (client side ON state) and the server responds with a RESPONSE (server side ON state). In between the requests and responses, both the entities have their OFF periods. The duration of the ON period depends on the sizes of the application layer messages. Typically, the REQUEST messages are small with a distribution showing limited variability whereas the RESPONSE messages are large with a distribution showing high variability. The application RESPONSE messages are fragmented into application layer buffers. These buffers are in turn broken up into packets (so-called ”cascading effect” [9]) to be carried by the specific network media being used. The high variability in the RESPONSE sizes directly results in the high variability in the packet arrival process. The OFF periods are typically due to user inactivities and the variability in the OFF process may be due to many different types of users accessing network services concurrently. The level of activity of human users generally varies from user to user. There may even be ”users” which are machines (HTTP proxy servers sitting behind a firewall or robots downloading Web pages to create mirror sites).

Transporting packets across the network between the application end points is subject to delays and errors. They may even be discarded by a congested router/switch. Because of the presence of LRD phenomenon across many types of networks, metrics of network performance such as throughput, packet loss, response time, buffer occupancy levels are affected [6, 16]. For instance, numerical and analytical studies of LRD traffic show that the tail of queue length distribution decays at a rate slower than that of the characteristic exponential decay as seen in traditional traffic studies. This deviation in the queue length process results in longer waiting times at the network processing elements (switches, routers etc). Furthermore, studies in [12] have shown the variation of some parameters (related to LRD) with link level load. Using a *fBm* model framework, the robustness of the Hurst effect has been shown to be an invariant under changing load conditions. This questions the effectiveness of traffic rate control mechanisms in altering auto-correlation structures seen in traffic streams. Other important consequences are:

- Packet delays and consequently application level delays have a heavy-tailed distribution.
- Transport layer protocols like TCP estimate the round trip timer values from the peer acknowledgements and hence are influenced by it.
- Congestion situations are unavoidable and they appear as short-lived impulses. With increase in load, congestions appear more frequently while maintaining the impulsive behavior.
- Only increasing buffer sizes does not result in significant improvement in packet loss behavior.

The main interest is in evaluating the consequences of LRD in the application layer quality on an end-to-end basis. The end-to-end delivery is the responsibility of the transport layer which encapsulates the manifestations of LRD property at the packet level traffic management. The transport layer characteristics are in turn modulated by the events occuring at the application layer and the way these events are managed at the application layer. It is important to understand and to quantify the effect of network controls (e.g, TCP flow control) on application-level parameters since the management of the application traffic has requirements that are different from those for the management of transport-level traffic. Application-level traffic management occurs by decisions on the admission control and resource management hierarchy whereas transport and network-level management requires the delivery of implied (e.g., IP) or contracted (e.g., ATM) QoS. Thus, in delivering end-to-end guaranteed service it is not just important to understand the traffic parameters at the application layer, but also to model and to quantify the effects of network control mechanisms on the application-level parameters. For example, if the application happens to be delivering multimedia service, then a stream of video and audio may be delivered to the client where the parameters of the video and audio stream are impacted by network congestion and recovery mechanisms. By the time the stream arrives at the client node, it may have been shaped significantly and the parameters of the stream may have changed drastically from the actual parameters negotiated during client-server connection setup. The case of interactive multimedia communications is more challenging as the source itself will shape the stream based upon feedback from the client. Thus, the issue of effective resource and traffic engineering to provide QoS on an end-to-end basis is a difficult question and is subject of ongoing and future research.

8 Conclusions

The traffic self-similarity existent in network and application traffic seems to be an ubiquitous phenomenon that is independent of technology, protocol and environment. As a consequence, many of the basic assumptions made in traffic control, engineering and operations and maintainance are questioned. This has also a profound impact on the performance, just to mention the potentially heavier queueing asymptotics than predicted by traditional Markov models. The network performance is dominated by the self-similarity property in the network traffic. Other important areas that are impacted by self-similarity are data analysis, statistical inference, mathematical modeling, queueing and performance analysis. These are questions that are still under investigation and research, and more efforts must be done in the future to answer them.

References

[1] Arlitt M.F. and Williamson C.L., *Web Server Workload Characterization: The Search for Invariants (Extended Version)*, IEEE/ACM Transactions on Networking, Vol. 5, No. 5, October 1997.

[2] Beran J., Sherman R., Taqqu M.S. and Willinger, W., *Long-Range Dependence in Variable-Bit Rate Video Traffic*, IEEE Transactions on Communications, Vol. 43, No. 2/3/4, February/March/April 1995.

[3] Crovella M.E. and Bestavros A., *Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes*, Proceedings of ACM SIGMETRICS'96, 1996.

[4] Causey J.W. and Kim H.S., *Comparison of Admission Control Schemes in ATM Networks*, International Journal of Communication Systems, Vol. 8, 1995.

[5] Duffy D.E., McIntosh A.A., Rosenstein M. and Willinger, W., *Statistical Analysis of CCSN/SS7 Traffic Data from Working Subnetworks*, IEEE Journal on Selected Areas in Communications, Vol. 12, No. 3, 1994.

[6] Erramilli A., Narayan O. and Willinger W., *Experimental Queueing Analysis with Long-Range Dependent Packet Traffic*, IEEE/ACM Transactions on Networking, Vol. 4, No. 2, April 1996.

[7] Erramilli A., Willinger W. and Wang J.L., *Modeling and Management of Self-Similar Traffic Flows in High-Speed Networks*, Network Systems Design, Gordon and Breach Science Publishers, 1999.

[8] Flandrin P., *Wavelet Analysis and Synthesis of Fractional Brownian Motion*, IEEE Transactions on Information Theory, Vol. 38, No. 2, March 1992.

[9] Feldmann A., Gilbert A.C., Willinger W. and Kurtz T.G., *The Changing Nature of Network Traffic: Scaling Phenomena*, Computer Communications Review, Vol. 28, No. 2, April 1998.

[10] Garett M. and Willinger W., *Analysis, Modeling and Generation of Self-Similar VBR Video Traffic*, Proceedings of ACM SIGCOMM'94, 1994.

[11] Jena A.K., Popescu A. and Pruthi P., *Modeling and Evaluation of Internet Applications*, submitted to IEEE/ACM Transactions on Networking, August 2000.

[12] Jena A.K., Pruthi P. and Popescu A., *Modeling and Evaluation of Network Applications and Services*, Proceedings of the RVK'99 Conference, Ronneby, Sweden, June 1999.

[13] Li G-L. and Dowd W.D., *An Analysis of Network Performance Degradation Induced by Workload Fluctuations*, IEEE/ACM Transactions on Networking, Vol. 3, No. 4, August 1995.

[14] Leland W., Taqqu M., Willinger W. and Wilson D., *On the Self-Similar Nature of Ethernet Traffic*, Proceedings of ACM SIGCOMM'93, 1993.

[15] Leland W., Taqqu M., Willinger W. and Wilson D., *On the Self-Similar Nature of Ethernet Traffic (Extended Version)*, IEEE/ACM Transactions on Networking, Vol. 2, No. 1, February 1994.

[16] Norros I., *A storage model with self-similar input*, Queueing Systems Theory and Applications, Vol. 16, Iss. 3-4, 1994.

[17] Park K., Kim G.T. and Crovella M.E., *On the Relationship Between File Sizes, Transport Protocols, and Self-Similar Network Traffic*, preprint, Boston University, 1996.

[18] Paxson V. and Floyd S., *Wide Area Traffic: The Failure of Poisson Modeling*, Proceedings of ACM SIGCOMM'94, 1994.

[19] Paxson V. and Floyd S., *Wide Area Traffic: The Failure of Poisson Modeling*, IEEE/ACM Transactions on Networking, Vol. 3, No. 3, June 1995.

[20] Peitgen H-O., Jürgens H. and Saupe D., *Chaos and Fractals New Frontiers of Science*, Springer-Verlag, 1992.

[21] Tuan T. and Park K., *Congestion Control for Self-Similar Network Traffic*, preprint, Purdue University, 1998.

[22] Taqqu M.S., Willinger W. and Sherman R., *Proof of a Fundamental Result in Self-Similar Traffic Modeling*, Computer Communication Review, Vol. 27, No. 2, April 1997.

[23] West B.J. and Deering B., *The Lure of Modern Science Fractal Thinking*, Studies of Nonlinear Phenomena in Life Sciences, No. 3, World Scientific Publishing Co., 1995.

[24] Wang J.L. and Erramilli A., *Monitoring Packet Traffic Levels*, Proceedings of GLOBECOM'94, 1994.

[25] Willinger W., Taqqu M.S., Sherman R. and 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, February 1997.