A decade of internet research: advances in models and practices

Carla Di Cairano-Gilfedder,
BT Group Chief Technology Office, Security Research Centre,
Flr 2 pp 5, Antares Building,
Adastral Park, Martlesham Heath,
Ipswich, Suffolk IP5 3RE, UK
Tel:+44(0)1473-640880, Email:carla.dicairano-gilfedder@bt.com

Richard G. Clegg,
Department of Mathematics,
University of York, YO10 5DD
Tel:+44(0)1904 433088, Email:richard@richardclegg.org

September 7, 2005

Abstract

The complexity and heterogeneity of the current internet have rendered traditional analytical models and techniques inadequate for networking researchers and engineers. Many within the networking research community feel that researchers investigating new protocols and architectures, either by simulation or test-bed implementation, need to use common models. Despite the lack of models with universal applicability, there are certain models that are more appropriate than others to analyse certain systems. Researchers in different areas tend to use established models, typically to allow comparison of results. In addition researchers have made considerable progress in understanding the statistical nature of internet traffic. Despite the widespread use of simulation and test-bed implementation for modelling the internet there is today little consensus on analysis techniques and their validation. This paper summarises the main advances made in the last ten years or so in understanding the nature of internet traffic, models and practices developed for internet topology and protocol dynamics analysis.

1 Introduction

Networking researchers are increasingly reliant on simulation and test-bed implementation to investigate the behaviour of new protocols as well as for the design and deployment of complex and heterogeneous networks. This complexity and diversity is constantly growing with the introduction of more end-users, introduction of new applications, and the deployment of wired and wireless technologies in large scale. The internet protocol suite is often used over these diverse networks to provide a platform for distributed applications and services. The size and complexity of these networks means that increasingly refined analytical and simulation tools are necessary to investigate their behaviour. Many within the networking research community [27] advocate that researchers should use common, realistic models. However, no model exists which can be applied to every possible networking problem. Nevertheless there are certain models that are more adequate than others to analyse certain systems. For example congestion control typically is studied with a single bottleneck topology, whereas multicast, peer-to-peer (P2P), and distributed denial of service (DDoS) attack studies necessarily require bigger topologies. Researchers in different areas use established models that most likely others have used before, often to facilitate comparison of results. In addition there has been considerable progress in understanding the statistical nature of internet traffic. This document provides a short review of the main achievements in this area over recent years.

The reasons for utilising internet models are manifold. The researcher may for example seek to study new protocols behaviour or seek to support the engineering processes of network dimensioning and provisioning. Whatever the motivations, it is crucial to identify the appropriate time-scale at which to analyse the problem at hand. This is because certain factors only exhibit themselves and therefore influence the system at certain time-scales. At the shortest timescales, ranging from microseconds to minutes, the influencing factors include protocol interactions with hardware and human factors such as for example how often a user is active. At medium timescales, with times ranging from hours to days, different factors become important, usually with a strong impact made by human behaviours which are influenced by time of day and time of year. Finally at the longest timescales, ranging from weeks to years, the behaviour of the network and its traffic are affected by seasonal patterns, long-term growth, and pattern changes due to the emergence of new applications. In particular, a change in the dominant application on the internet may radically reshape overall traffic characteristics. Notably, in the late nineties network traffic was largely generated by web based applications; as of 2005 the dominant position is occupied by peer-to-peer traffic which exhibits different characteristics than HTTP. In this paper only the shortest timescale, ranging from microseconds to minutes, will be treated. It is in fact at these timescales that engineering factors have stronger impact than human behaviours and network economics.

This document is organised as follows. In Section 2 a treatment is given to current understanding of the statistical nature of internet traffic. Section 3 discusses issues connected to internet topology modelling. Section 4 concludes the document by summarising current consensus, known pitfalls and guidelines for modelling the internet.

2 Internet traffic characteristics

In order to successfully simulate internet traffic or capture it in a mathematical model, an understanding of its statistical properties is necessary. Internet traffic has been shown to exhibit correlation over a range of timescales in different networking contexts with potentially detrimental effects on performance. This section gives an overview of work that in the last decade has provided evidence of existence and found causes of traffic correlation and discusses models that have been developed to study networks with correlated traffic.

Before the early 1990s traffic and performance studies had been predominantly based on models such as Poisson processes which have no long-term correlation structure. Such models are attractive because of their mathematical tractability and the large body of queuing theory which relies on the assumption of Poisson processes. In 1993, the seminal paper [39] (expanded in [40]) found evidence of long-range correlation in LAN traffic and brought the concept of self-similarity (and the related concept of long-range dependence) into the field of network traffic and performance analysis. As a result, it is often considered important to capture these correlation aspects when modelling internet traffic. The concept of self-similarity has a long history in mathematics but Mandelbrot was an important pioneer [41]. Self-similarity is informally described as the property of an object which looks “the same” when viewed at different scales. For network traffic this can be thought of loosely as the idea that the traffic is bursty in the same way whether the time-scale we are considering is milliseconds or seconds (the approximation being only valid over a limited range). A more formal definition will be given later where we will also see that traffic self-similarity has been observed in several contexts: LAN and WAN networks, IP and ATM protocol stacks.

2.1 Basic statistical concepts for network traffic

In this section we rigorously define the three statistical properties of network traffic of *long-range dependence* (LRD), *statistically self-similarity* and *heavy-tails.* These properties are inter-related and refer to traffic scaling behaviour that is they are measures of how a system looks when considered at different scales.

Let $\{X_1, X_2, \dots\}$ be an infinite time series. A time series is said to be *weakly-stationary* if it has a constant and finite mean ($\mathrm{E}[X_i] = \mu$ for all i, where E means expectation) and the covariance between X_i and X_j (i.e. $\mathrm{E}[(X_i - \mu)(X_j - \mu)]$) depends only on $|j - i|$.

If the time series is weakly-stationary then the autocorrelation function (ACF) $\rho(k)$ is given by

$$\rho(k) = \frac{\mathrm{E}[(X_t - \mu)(X_{t+k} - \mu)]}{\sigma^2},$$

where μ is the mean and σ^2 is the variance.

An important statistical property of network traffic is that of *long-range dependence*, also known as *long memory* or sometimes *strong dependence.* Several definitions (not all equivalent) exist in the literature. For a good book on the topic see [8] and for a beginners summary in the context of telecommunications see [12, chapter one]. A commonly used definition is as follows:

Definition 1. A weakly-stationary time series $\{X_1, X_2, \dots\}$ is said to be *long-range dependent* if the sum of the autocorrelation diverges (that is $\sum_{k=0}^{\infty} \rho(k)$ is not finite).

Often a specific functional form for the autocorrelation for a LRD process is assumed, that is

$$\rho(k) \sim C_\rho k^{-\alpha}, \tag{1}$$

where $\sim$ means asymptotically proportional to (that is $f(x) \sim g(x)$ means $\lim_{x\to\infty} f(x)/g(x) = 1$; in certain cases the limit may be as $x \to 0$) and where $C_\rho > 0$ and $\alpha \in (0, 1)$. The best known measure of LRD is the Hurst parameter H where $H \in (1/2, 1)$ indicates LRD, $H = 1/2$ indicates independent or short-range dependent data and as H grows towards $H = 1$ stronger degrees of LRD arise. The parameter α is related to H by the equation $H = 1 - \alpha/2$.

Sometimes, it is useful to consider LRD from the point of view of the frequency domain. Consider the *spectral density* of a weakly stationary time-series. This can be related to the ACF via the Wiener-Khinchine theorem [71]. If $f(\lambda)$ is the spectral density of a time series at frequency λ then

$$f(\lambda) = \frac{\sigma^2}{2\pi} \sum_{k=-\infty}^{\infty} \rho(k) e^{ik\lambda},$$

where $i = \sqrt{-1}$, $\rho(k)$ is the ACF of the series and σ^2 is the variance of the series.

In the frequency domain equation (1) becomes

$$f(\lambda) \sim c_f |\lambda|^{-\beta},$$

as $\lambda \to 0$, where $f(\lambda)$ is the spectral density, $c_f > 0$ and $\beta \in (0, 1)$. The parameter β is related to the Hurst parameter by $H = (1 + \beta)/2$.

From these definitions, the concept of LRD can be considered in two separate but related manners. Firstly, LRD can be thought of as correlations persisting over an extremely long period of time. Secondly, it can be thought of as a significant power in low frequency bands.

A related concept is that of self-similarity, this is defined as follows.

Definition 2. Let Y_t be a stochastic process with continuous time parameter t. If the process is *self-similar* with self-similarity parameter H then for any positive constant c, the rescaled process $c^{-H} Y_{ct}$ is equal in distribution to the original process Y_t.

This definition, in essence, says that if a process is self-similar, then it looks "the same" in a statistical sense when considered at different time-scales, in other words if the x (time) axis is stretched then the process will have the same distribution if the y axis is also stretched by an appropriate factor.

If a self-similar process Y_t has stationary increments and $H \in (0, 1)$ then it can be shown (see [8, page 51]) that the increment process given by $X_i = Y_i - Y_{i-1}$ for $i \in \mathbb{N}$ has an ACF given by

$$\rho(k) \sim H(2H - 1)k^{2H-2},$$

which implies that for $H \in (1/2, 1)$ then the increment process is long-range dependent. This gives the relationship between self-similar and long-range dependent processes.

Another important statistical concept is that of heavy-tails. A variable is said to be *heavy-tailed* distributed if the tail of the distribution function decreases to zero more slowly than exponentially. Formally, for all $\varepsilon > 0$, a random variable X is heavy-tailed if it satisfies

$$\mathbb{P}[X > x] e^{\varepsilon x} \to \infty, \text{ as } x \to \infty.$$

Often a specific functional form is assumed for the distribution:

$$\mathbb{P}[X > x] \sim Cx{-}\alpha,$$

for some $C > 0$ and some $\alpha \in (0, 2)$. Intuitively a heavy-tailed distribution implies that large values can occur with a non-negligible probability. It has been shown [66] that a superposition of ON/OFF processes where the lengths of ON and OFF periods are heavy-tailed will give rise to a self-similar process. It can also be seen that heavy-tails in the ON and OFF periods could lead to long-range correlations in the process itself.

A widely known heavy-tailed distribution is the Pareto distribution given by

$$\mathbb{P}[X > x] = bx^{-\alpha}, \text{ for all } x > b,$$

where $b > 0$ is the minimum value of X and $\alpha \in (0, 2)$ in order to be a heavy-tailed distribution.

The presence of power-law relationships imply that processes which are long-range dependent or self-similar and distributions which are heavy-tailed, in some sense look the same when viewed at different scales. A process which scales in a constant way is sometimes referred to as *mono-fractal*. A generalisation of this is a *multi-fractal* process which exhibits complex behaviour that changes over different timescales [46]. As we will see later, network traffic at very small timescales, of the order of few hundred milliseconds, may be better modelled by means of *multi-fractal* processes.

The issue of measuring traffic LRD and statistical self-similarity is a complex one. In the time domain it is characterised by the fall off of the ACF at high lag. However, these are exactly those areas at which the fewest readings are available and the data is most unreliable. Similarly, in the frequency domain, the LRD is characterised by the behaviour of the spectrum at frequencies near zero, which are precisely those frequencies which are hardest to measure. It is certain that simply examining the ACF is not a robust way to estimate the Hurst parameter. In addition, a number of biases may be present in real-life data which could cause problems. These include periodicity (users and processes daily usage patterns) and trends (traffic volume changes throughout the measurement period) which violate the assumption of weak-stationarity and create difficulties for estimations. The topic of measuring LRD is beyond the scope of this paper, the reader is referred to [65, 6] for work which compares existing techniques. Also a summary paper giving practical advice on the estimation of the Hurst parameter in real data and caveats about the associated pitfalls is [13].

2.2 Internet traffic LRD: empirical evidence

The paper [39] established for the first time that network packet traffic has self-similar characteristics. Evidence of self-similarity was found by means of a rigorous statistical analysis of a large set of LAN (Ethernet) packet traces taken at Bellcore. This paper is of considerable importance as, for the first time, accuracy and applicability of existing (mainly Poisson) models for the analysis of networks performance were questioned. In the light of this paper, Floyd and Paxson [52] embarked into a study aiming at showing that WAN network traffic does not exhibit self-similarity properties. To their surprise they found that WAN traffic is consistent with self-similar scaling. The existence of self-similarity in WANs was later confirmed in [17] and [24]. In [52] it is also shown that Poisson processes adequately model certain session arrivals such as FTP and TELNET but not others such as HTTP, SMTP and NNTP. Evidence of self-similarity in WWW traffic was provided in [17]. The authors use a large set of WWW traces taken in a university computer lab over several months. They show that self-similarity in Web traffic could be explained by the long-tailed nature of the distribution of file sizes, the effect of user behaviour, and the aggregation of multiple flows in local area networks.

These early papers which found evidence and explanations for self-similarity in network traffic, concluded that self-similarity is an intrinsic property of traffic generated according to certain distributions of session duration and size, regardless of how packets are sent to the network within each session. Later research, focusing on WAN network traces, questioned at which level the network impacts on traffic characteristics, i.e. when it matters how the packets are sent within each session. In [56] and [24] it was found that the impact of the network (e.g. protocols, topology) shows up when studying the network at small time-scales, i.e. timescales of the order of the RTTs in the system investigated, typically then of the order of a few hundred milliseconds or smaller. At these small timescales traffic exhibits a structure more complex than self-similar, namely a structure that is consistent with multi-fractal scaling. These findings did not invalidate earlier work but complemented it, in that they established that at small timescales internet traffic behaviour is consistent with multi-fractal scaling whereas at larger timescales it is consistent with self-similar scaling.

Whereas earlier work had shown that the aggregation of multiple flows leads to traffic exhibiting self-similar properties, in [31] evidence was found that VBR video traffic deriving from a single flow is long-range dependent.

In summary evidence has been found that:

- Self-similarity is exhibited by LAN (Ethernet) and WAN (internet) traffic at timescales longer than a few hundred milliseconds.
- In a WAN environment a connection starts at a random point in time, begins transmitting and then stops. The connection arrival process can be modelled by means of a Poisson process.
- By choosing the distribution of lengths of ON/OFF periods (for the LAN case) or the distribution of the session durations (for the WAN case) to be heavy-tailed with infinite variance, one can observe that aggregated packet level traffic exhibits asymptotic self-similarity behaviour over timescales of the order of few hundreds milliseconds or larger.
- Multi-fractal scaling is exhibited by LAN (Ethernet) and WAN (internet) traffic at timescales of the order of the RTTs in the system under consideration, typically of the order of a few hundred milliseconds.

2.3 Internet traffic LRD: possible causes

Several studies have been carried out to explain the physical causes of self-similarity in network traffic and identify the properties of distributed systems that induce self-similarity at multiplexing point. In [47] the so-called self-similarity "structural causality" was discussed: if the distribution of file or object sizes is heavy-tailed then, at multiplexing points, network layer traffic is self-similar. This result was shown by simulation and found to be consistent over a range of link bandwidths and buffer sizes. We have seen earlier that, intuitively a distribution is heavy-tailed when it can take a wide range of values and very large values may occur with non-negligible probability. Heavy tails have been observed for file sizes on file and web servers and CPU times of processes, as reported in [47]. The reason why heavy-tails cause self-similarity was discussed in [16] and [72]. These papers show that multiplexing a high number of independent ON/OFF sources with heavy-tailed strictly alternating ON and/or OFF periods gives rise to self-similarity. In [23] it is shown that the heavy-tailed nature of file sizes at the application layer is preserved down the protocol stack and leads to approximate heavy-tailed busy periods of network layer traffic. In [23] it is also proved that self-similarity over large time-scales is almost entirely attributable to user related variability, as opposed to being network (e.g. protocol, topology) related. Hence packet level self-similarity reflects the high-variability at the session level. The breakpoint between small and large time scale occurs at some intermediate point that depends upon RTTs for TCP or rate for UDP, it also depends upon variability of congestion and delays. At short time scales the more complex behaviour of multi-fractal scaling is observed (as opposed to self-similar scaling). A convincing explanation for this behaviour at short-time scales has not been found yet. It has been hypothesised that this is caused by the presence of TCP-like flow control algorithms and, to a lesser extent, due to network-related variability. For example, [26] uses a Markov approach to model TCP timeout behaviour and shows that it can lead to what the authors refer to as "local long-range dependence" (that is, LRD up to a certain time scale). However, this has been called into question by recent work. Specifically, some of the original measurements regarding the TCP mechanism have been cast into question [25] and shuffling experiments have shown that LRD doesn't disappear if artifacts due to the TCP feedback mechanism are removed from flow patterns [34]. Hence, the claim that scaling behaviour in internet traffic arises from TCP feedback mechanisms must now be regarded as doubtful.

Finally, there remains the possibility that LRD is an emergent property from the network itself. Measurements made in [11] show that, even when "packet inter-departure times are independent, arrival times at the destination show LRD". Simulation work has shown that in simple simulations where all sources are Poisson and no TCP-like feedback mechanisms are used the emergent traffic can exhibit LRD [63, 74].

In summary, therefore, four possible causes for LRD are identified in the literature:

- Some traffic such as VBR video traffic intrinsically pos-

sesses LRD because of its nature.
• The aggregation of file transfers where the sizes of the files has a heavy-tailed distribution leads to LRD.
• TCP feedback mechanisms lead to LRD. However, note that this claim must now be regarded as doubtful.
• LRD arises from the network itself as an emergent property.

2.4 Traffic correlation and control

In the previous section we have seen that self-similarity has been explained as an intrinsic property of traffic generated at the application layer. We have also observed that evidence has been found that the network itself gives rise or contributes to self-similarity. The paper [47] argues from simulation results, that the flow control mechanism and reliable transmission of TCP (Reno, Tahoe, Vegas) preserve traffic self-similarity at the link layer at various timescales. In these simulations, the presence of closed-loop control results in a direct relationship between the heavy-tailed distribution of transmitted data files and self-similarity at the link level traffic. In contrast, in an open-loop simulation based on UDP, heavy-tailed distribution of file size is seen to generate bursty traffic at short timescales, but does not generate long-range dependence. This result was shown to be robust to changing in file distribution, topologies, link capacities, buffer sizes, and interference from different cross traffic.

The papers [56] and [24], mentioned earlier, claim that at time-scales of the order of a few hundred milliseconds or smaller network traffic is consistent with multi-fractal scaling. This could be caused by the presence of TCP-like flow control algorithms (however see [34] for a counter-claim) and, to a lesser extent, due to network-related variability. Research effort is currently being put into deriving further multi-fractal models and finding a physical explanation for multi-fractal behaviour at small time-scales [20].

In [53], by means of simulation, it is argued that TCP retransmission mechanism alone may give rise to self-similarity at various time-scales. More recently the paper [68] claims that TCP congestion control alone can cause self-similarity regardless of application layer traffic characteristics. However, these measurements are contested by [25]. The impact of TCP control on traffic self-similarity has been investigated more recently in [73] by means of simulation. Poisson arrivals are assumed at the application layer and results show that traffic in the (bottleneck) network looks Poisson when either the network is lightly utilised with little loss or when the network is highly utilised, in these circumstances TCP congestion control smooths out the burstiness of the aggregated traffic. However, the traffic appears self-similar when the network is not stable with intermediate load and occasional short periods of congestion (typical of many networks).

It is hard yet to make clear conclusions on the exact effect of closed loop (TCP) interactions on traffic correlations. The evidence currently seems inconclusive. Simulation studies and mathematical models show that closed loop interactions can lead to long-range correlations in traffic. However, experiments on real data show that when the data is shuffled to remove the effects of closed loop interactions then LRD is still present although the nature of correlations does change. It seems that more research is needed in this area before firm conclusions can be drawn about what effect TCP feedback mechanisms do genuinely have on real networks.

In summary, findings regarding relationship between traffic correlation and control are:
• In a network, even without the presence of feedback mechanisms, the heavy-tailed distribution of file sizes may introduce long-range dependence.
• TCP retransmission and congestion control mechanisms have been suggested as causes of long-range dependence but this remains controversial.
• In a network where traffic levels are low the degree of long-range dependence is also low (the Hurst parameter is near 0.5) and SRD models may be appropriate. More congestion can lead to a higher Hurst parameter up until the point where the network becomes overloaded.

2.5 Traffic correlations and queuing theory

For the performance analysis of queues with correlated traffic input, new queuing models have been developed [1], [2], [44], [48], [67]. These models apply to infinite buffers and suffer from their asymptotic nature, that is the theoretical expectation that they provide are obtained when buffer length is either assumed to be infinite or tends to infinity. Under the assumption of infinite length buffer and long-range dependent traffic input the main finding is that the distribution of queue length has slower than exponential decaying tail (as opposed to exponential for traffic with non-long-range dependence). The queue length decaying function has been described by a Weibull distribution [44] and by polynomial [67]. If the arrival process is described by a single ON/OFF source with heavy tailed ON and OFF periods than the queue length distribution decays according to a power-law. Whereas if the arrival process is described by a single ON/OFF source with OFF periods only being heavy tailed, than the queue length distributions decays exponentially [46].

Little is known about finite buffer systems where the traffic is highly correlated. Some authors question the importance of capturing traffic long-range dependence when finite buffer sizes are considered [32] and [62]. It is argued that correlation becomes irrelevant for small buffers and short time-scales, and hence suitable short-range dependent models are adequate for self-similar traffic modelling. In [62] it is shown that, for realistically configured ATM switches' buffer sizes, cell loss is not adversely effected by long-range dependent traffic and that short term correlation have a much more dominant effect on performance and therefore Markovian models are adequate for analysing performance.

We have seen earlier that in [47] it is shown that the degree to which file size distribution is heavy tailed (at the application layer) directly affects the degree of self-similarity at the link layer. They investigate the effects on queuing behaviour in the presence of self-similar traffic and in closed-loop system based on TCP. They find that queuing delay is strongly affected by an increase in the degree of self-similarity, whereas packet loss and re-transmission rate increase only gradually (roughly linear).

A key lesson learnt is that, from a queuing dynamics perspective, multiplexing is highly desirable to smooth out traffic burstiness, as the addition of independent flows reduces variances and the impact on queue lengths deriving from traffic bursts.

Initial efforts in understanding effects of self-similarity have concentrated on first order performance measures such as loss and delay. Further research effort needs to be put into understanding transient queuing behaviour under self-similar traffic input as well as understanding second-order performance measures such as jitter and packet loss variation (particularly relevant when provisioning services requiring certain QoS guarantees).

In summary the main findings regarding relationship between traffic correlation and queuing theory are:
• In infinite buffer systems the queue length has been shown to delay slower than exponentially.
• Further research needs to put into understanding queueing dynamics in finite buffer systems.
• Further research needs to be done towards better understanding of second-order performance measures (e.g. jitter, loss variation).

2.6 Self-similar traffic models

The analysis of long-range correlation behaviour in the context of network simulation can be based on a number of models. For a comprehensive review of these models the reader is referred to [46]. In choosing a model there is often a trade-off between its simplicity and its ability to capture the nature of the traffic. A few of the more commonly used models are listed in this section.

Fractional Brownian motion (fBm) is a non-stationary stochastic process which is a generalisation of the well-known Brownian motion (or Weiner process) but with a dependence term between samples. It is a self-similar process and has a defined Hurst parameter. (If $H = 1/2$ then it is simply Brownian motion). If $B_H(t)$ is fBm then the process $Y_k(t) = B_H(t+k) - B_H(t)$ when $H \in (1/2, 1)$ is *fractional Gaussian noise* (fGn) which is long-range dependent. A quick computational method for generating fGn is given by [50]. This is an extremely simple model which only has two parameters: the Hurst parameter and a variance parameter. This makes it mathematically attractive but its simplicity means that it cannot capture a diversity of mathematical properties. In addition, by definition the fBm model must at time produce negative results (albeit with a small probability). However a traffic model that predicts negative packet arrivals or negative inter-arrival times might be considered flawed.

Fractional Auto-Regressive Integrated Moving Average (FARIMA) models are an expansion of the classic time-series ARIMA models and allows modelling of long-range dependence. In addition to allowing the specification of a Hurst parameter, they can also allow the specification of degrees of long-range dependence. A description in the context of LRD can be found in [8, pages 59–66]. However, as the fGn seen before, they also suffer from the drawback about possible negative inter-arrival times or packet arrivals.

LRD can be also be generated using a family of chaotic maps known as intermittency maps. For example take a map from the family given by

$$x_{n+1} = \begin{cases} x_n + \frac{1-d}{d^{m_1}} x_n^{m_1} & 0 < x_n < d, \\ x_n - \frac{d}{(1-d)^{m_2}} (1-x_n)^{m_2} & d < x_n < 1, \end{cases} \quad (2)$$

where $d \in (0, 1)$ and $m_1, m_2 \in (3/2, 2)$. This map can be used to generate a binary time series which is one if $x_n < d$ and zero otherwise where the ones and zeros are interpreted as packet transmit and inter-packet transmit gaps. If $m = \max(m_1, m_2)$ then this will generate LRD traffic with $H = (3m - 4)/(2m - 2)$. The map is analytically difficult to work with and hence piecewise linear approximations have been studied using Markov chains. Pioneering work in this area is [69] with early applications to telecoms being given by [21]. Two different approaches to using Markov chains for generating LRD are given by [7, 14].

An interesting technique for modelling traffic is Wavelet analysis. This allows not only capturing the Hurst parameter but also the synthesis of a whole host of scaling behaviour and the replication of the multi-fractal spectrum. Details can be found in [58, 57].

There is no consensus on which of the proposed models is the most adequate to use; however when synthesising network traffic for simulation studies or analysing traffic measurements, some maintain a useful rule-of-thumb is to use as many different methods as possible for checking and validating whether or not the data at hand is consistent with the hypothesised scaling behaviour.

Analysing sophisticated protocols under correlated traffic conditions is a complex task as often protocols and their control mechanisms have an impact on the traffic itself. As discussed in [45], incorporating correlation in models for protocol analysis and simulation can follow two approaches:

- **Trace-based approach**: In this approach the aim is to produce traffic as statistically "close" to the real traffic traces as possible. Real traffic traces from a variety of sources can be analysed and mathematical or computational models can be developed which generate traffic which is similar in statistical nature. The wavelet based method previously described is a good example of this.

- **Network-based approach**: In this approach the aim is to produce as good a simulation of the internet as possible with the hope that this will automatically generate traffic with the required statistical nature. By an accurate simulation of protocols and network topology it is hoped that the statistical nature of the traffic will show up an emergent property.

Both approaches have associated drawbacks. With the trace-based approach, traffic can be produced and used to answer questions about queuing behaviour and about network capacity requirements. However, a fundamental problem with this approach is that the internet relies on protocols that control congestion and feedback. If the network changes the statistical nature of the traffic changes with it. Indeed the statistical nature of the traffic is a fundamental property of the protocols *and* the topology of the network. On the other hand, with the network-based approach, it is impossible to accurately simulate a network as vast and complex (and ever-changing) as the internet. A network-based approach is therefore unlikely to ever be able to fully capture the dynamics of the internet.

3 Modelling internet topology

The topology used to model network behaviour under different protocols and loads needs to be as representative as possible of the conditions arising in the real network and needs to allow comparison between different protocol studies. While ideally the performance of protocols should not be affected by topology, it has been found that often protocols behave differently in different topologies. The topology that is most commonly used for modelling congestion control is the *dumb-bell* (also known as the *barbell*) which includes a single bottleneck link with multiple transmitting and receiving entities. Another topology used consists of a single path but with multiple bottleneck links, see Floyd [28]. Keshav in [38] attempts to model a wide range of handcrafted network topologies and loads. In addition to these simple and intuitive models, topologies have often been generated by means of automated traffic generators. The first network topology generator that was used for protocol simulations was proposed by Waxman [70]. This randomly generates links with a probability that depends upon the (assumed) distance between nodes. It was later argued that real network topologies do not have a random structure and that real networks' modelling must account for networks hierarchical structure. As a result, a family of network generators, called *structural* generators, which focused on the hierarchical nature of networks was developed; see for example [19]. This set of generators dominated the scene until Faloutsos et al.[22] showed that routers in internet graphs have a degree (i.e. number of attached links) distribution that is heavy-tailed. Because of this connectivity propriety these networks are often referred to as *scale-free*, as they self-organise into a scale-free state. These findings had a great impact on the topology generation philosophy at the time, as structural generators do not produce topologies with heavy-tailed node degree distribution. Researchers then focused on developing topology generators that matched the measured internet nodes degree distribution (e.g. [37], [4], and [3]). This family of generators, known as *degree* or *measurement* based, have been extensively adopted since. For networks with a medium to high number of nodes (~ 1000) in [64] it is maintained that, while focusing on local properties (i.e. of degree distribution) degree-based topology generators can capture the large-scale hierarchical structures of real networks. The heavy-tailed node degree distribution gives rise to the network hierarchical structure. However [76] points out that a power-law distribution is almost meaningless if the number of nodes is small. This means that, when modelling a smaller topology (< 100 nodes), degree-based generators will not be suitable as they would be unable to produce a hierarchical structure, therefore structural generators may still be preferable. In addition, degree-based generators have been the object of criticism by some who argue that power-laws are ubiquitously expected from a statistical point of view when dealing with high variability data. In addition BGP statistics that led to conclude about the existence of power-laws distributions have been proved to be incomplete. Also some maintain that degree-based topology generators are useful from a purely descriptive viewpoint but are incomplete models, as they are not driven by technical and economic considerations that in practice have considerable impact on network design. Research efforts in the field of network topology generation have therefore been put into designing topology generators that are more representative of real networks by incorporating both hierarchical and nodal degrees features of the actual internet topology [43].

It was noted earlier that the dumb-bell topology is extensively used for congestion control modelling. [5] argues that studies based on the dumb-bell topology may be misleading and hence conclusions drawn from its study must be considered with care. The authors argue that the dumb-bell topology is not representative of the internet as, in reality, there is a considerable proportion of traffic that traverses several congested links (at least two). Furthermore they argue that dumb-bell topology is not representative of the internet topology as this usually assumes that the congested link is a backbone link, but in practice backbone links are generally over provisioned and a bottleneck point is more likely to lie at the edge of backbones, at access points. Also, modelling a scenario with multiple bottlenecks highlight certain behaviour that would not be predictable with a dumb-bell topology. Indeed they show that a number of TCP flows with long round-trip times going through multiple congestion links get more bandwidth than TCP flows with short round-trip times traversing only one bottleneck link.

In [5] it also is maintained that no single topology or single family of topologies currently known is universally adequate for simulation studies. But the research community has a need to find a topology suite that is widely accepted, so as to allow comparisons between different protocol studies. In fact one of the reasons for dumb-bell's popularity has been the fact that others had used it before and hence results were comparable. Such a widely accepted topology suite will also need to be as simple as possible, as bigger topology may make interpretation of results complex and increase researchers' likelihood of drawing incorrect conclusions. It would seem that considering multiple topologies is advisable, although this would increase complexity of interpretation. The authors' recommendation is to use a carefully crafted network topology and loads that produce relevant behaviour that are known to be present in the real network and that may have generated abnormalities with other protocols (see Section 4 for more details).

In addition to a widely accepted topology suite, currently many in the research community feel the need for widely accepted network models. A network model includes topology but also traffic characteristics, congestion levels, protocols, scheduling policies etc. It is not clear, at present, whether the models being used are valid or flawed. In [30] Floyd and Kohler argue that it not conceivable to design a unique network model that can be employed in all circumstances and, at the same time, draw meaningful conclusions. Rather network models should be crafted by the researcher on the basis of the problem at hand so as to capture the relevant behaviour and ignore redundant ones. The Table 1 taken from [30] contains an overview of typical models that have been used in different contexts with assumptions that the researcher is required to make.

No tool is currently available to model and simulate a network of the size of the internet. Existing tools for the simulation of large-scale topologies are limited in the number of nodes that can be represented. Some of these tools are detailed in Table 2, as reported in [59].

Floyd and Paxson [51] argue that accurately simulating the internet is an impossible task as this is a "continuously moving target" affected by continuous topological, traffic, and routing changes. In [59] it is shown that, under fairly conservative assumptions regarding the number of internet hosts (∼110,000,000) and their traffic demands, simulating a network of the size of the internet is computationally not tractable neither with current computing capabilities, nor it is likely to be tractable in future. They observe that simulating a network of the size of the internet for 100 seconds would require more than a year of CPU time, nearly 300 Terabyte of memory, and about 1.4 Petabytes of disk storage to log the results. The authors therefore stress the importance of smaller scale simulations whose outcome can be extended with a sufficient degree of confidence to larger topologies. They suggest that a possible approach could be to start with a small simulation (with a few nodes) and then gradually increase the size of the network investigated. They provide examples of simulations where for topologies with an increasing number of nodes, "results" tend to converge. Of course they do not claim that this convergence will universally apply to all simulations, but this approach may be worthwhile in order to increase confidence in simulation findings. These authors also maintain the unsuitability of the dumb-bell topology, but they show (by experiment) that modelling larger topologies is not a trivial task and may lead to results that are difficult to interpret. More specifically, they compare the performance of the RED [29] and DropTail queue in a dumb-bell topology and in much larger topologies and found that, while clear difference existed between the performance of the two queues regime with the dumb-bell topology, with much larger topologies sometimes no difference was observed, while results sometimes suggested better performance with RED and sometimes with DropTail.

A simple alternative to the dumb-bell topology is the parking-lot topology [10] (see Figure 1). This has the advantage that it is not complex hence allows fast simulations and interpretation of results. This topology has five routers that are connected by the bottleneck links. This topology allows modelling flows traversing a different number of bottleneck links (i.e. 2, 3, and 5) and a variety of RTTs.

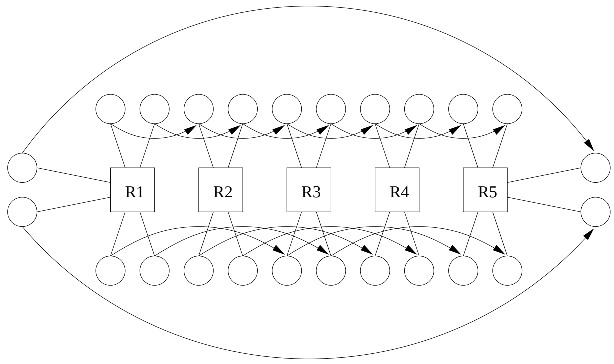

Figure 1: Parking lot topology

4 Modelling the internet: lessons learnt

Network models used in practice do not often resemble internet reality. This in itself is not a problem as by considering simple models in simulation or test-bed implementation one aims to gain insight into system behaviour in a simplified context. This section contains a set of guidelines that the network researcher may benefit

Research Topics	Typical model used	Supporting assumptions
AQM, scheduling, differentiated services	A dumb-bell topology, with aggregate traffic	Characteristics of congested links, range of RTTs, traffic characterisation, reverse-path traffic, effects of congestion elsewhere
Unicast congestion control	A single path, with competing traffic	Characteristics of links, queue management along path, variability of delay, bandwidth asymmetry
Multicast congestion control	A single multicast group in a large topology	Router-level topologies, loss patterns, traffic generation by group members
Routing protocols	A large topology	Router-level topologies, AS-level topologies, loss patterns
Routing lookups	A lookup trace, or a model of the address space	Ranges of addresses visible at a link
Web caching, peer-to-peer systems	Models of large topologies with application traffic	Topologies, application-level routing, traffic patterns
Controlling DDoS attacks	Models of large topologies with aggregate traffic	Topologies, attack patterns

Table 1: Commonly used simulation models and required assumptions.

Tool name	Reference	Number of modelled nodes
OPNET	[9]	Few hundreds
GloMoSim	[77]	Few thousands
NS	[42]	Few thousands
TeD	[54]	Few tens of thousands
pdns	[60]	Up to few hundred of thousands
SSF	[15]	Up to few hundred of thousands
USSF	[55]	Up to few hundred of thousands

Table 2: Popular simulation tools.

from when building and analysing a simplified model of the internet.

• When building and studying the behaviour of a system, researchers must understand how sensitive the system is to parameter changes. It is therefore necessary to explore a range of parameters and understand consequences and their causes.

• Care must be taken when analysing the sensitivity of a system to parameter changes. While modelling may show that the system is highly dependent to certain changes the internet might not. This is because a considerable impact of certain changes may be the result of artificially selected models.

• A rule of thumb to use when studying parameter sensitivity is to fix all parameters except one, and investigate the impact of changing this parameter over several orders of magnitude.

• Lessons learnt from the vast literature that followed the paper that introduced RED [29] highlighting RED weaknesses are that the researcher should bear in mind that overlooking the following issues may lead to wrong conclusions:

- How does the system behave when there is high packet drop rate ?
- How are conclusions affected by the size of the link(s) modelled ?
- How does the system behave if there is an oscillation of queue size ?

• Traffic mix can strongly affect queue length dynamics, for example adding short-lived flows, reverse path traffic, and different RTTs may change simple oscillations into more complicated bursty behaviour.

• When making assumptions about RTTs it is advisable to model a broad range of RTTs, including short ones ($<$ 50 ms) [36]. One may ask if the modelled range corresponds to what happens in practice. This can be verified by recording RTTs observed during the simulation and comparing results with real RTT distribution observed in the internet. If the distribution of the simulated RTTs is severely different from measurements, then transmission delays should be adjusted accordingly.

• By recording cumulative distribution of packet numbers sent during a simulation, it is possible to find out the fraction of flows in slow start (Web traffic) and the fraction of long lived flows at a particular time of the simulation and verify if this traffic mix is representative of real networks.

• When deciding upon parameters to model it is advisable to consider available up-to-date internet statistics.

• Models must be derived keeping in mind that protocols

or architectures are for the “Internet of the future” and not the one of today or the past.
• Researchers should avoid the pitfall of modelling packet-level trace-drive simulation where traffic derives from network measurements. This is because, on the internet, TCP feedback mechanisms will adjust the packet sending rates. Considering real network traces in an idealised network is not necessarily realistic [51]. Trace-driven simulation that attempt to replicate source-level behaviour provide a more plausible model. This is because the network affects how and when traffic reaches destination but not source behaviour [51]. Source-level behaviour was modelled for example in [18] and [49].
• On the internet virtually all levels of congestion are observed, hence a comprehensive analysis cannot dismiss any one in particular. [51].
• Modelling congestion should take into account the characteristics of congestion, for example, [35] observed that in the internet there are intermittent short-lived spikes of congestion.
• When comparing results one must take care on exactly which protocol implementation was used. For example performance achieved with TCP Reno may be very different than the one obtained with later TCP versions that corrected Reno's problems arising from multiple losses in one RTT.
• Aggregation of statistics such as throughput, arrivals, and duration of connections can be used to analyse the behaviour of a system. However, a careful aggregation process is necessary as inappropriate aggregation may give rise to misleading results. For example, protocols burstiness may be missed by comparing average data transmitted in a given interval [33].
• To increase confidence, simulation studies should be reproducible by others and hence models made available to the public.

4.1 Modelling traffic

• Studying TCP traffic requires models that take into account its long-range dependence. The use of Poisson models may underestimate performance measures such as queue size and packet delay. Poisson models can be used for certain types of session arrivals. Accurate modelling of wide-area traffic should be based on self-similar traffic models.
• Aggregate internet packet arrivals exhibit long-range dependence and are described by self-similar processes. Several methods have been proposed to generate self-similar traffic. However, there is not yet agreement of the most suitable self-similar synthesiser. The ON/OFF model combines link level self-similarity with source level behaviour.
• In 2001 about 90-95% of internet traffic was carried by TCP (www pages, data files, MP3 tracks). This means that the vast majority of traffic was elastic with changes dictated by TCP congestion avoidance scheme. A small fraction, but increasing, involved inelastic streaming audio and video. So modelling the internet of the future should account for this trend.
• The arrival process of flows at a backbone link typically is given by the combination of a large number of independent sessions. This arrival process can be described by a Poisson process.
• Long-range dependence and high variability may sometimes cause traffic starvation. This can happen when a link applies priority queuing and no bandwidth restrictions exist for the higher priority traffic. In this case, high priority traffic that has a long-range dependence and high variability over long time scales may starve low priority traffic.
• Incorporation of self-similarity in models for protocol analysis and simulation can follow the two approaches: network-based and trace-based approaches discussed in Section 2.6.
• Users session arrivals are well described by a Poisson process, provided that the hourly rate of this process is allowed to change in order to account for daily or weekly patterns [52].
• [49], [52], and [75] suggest the breakdown in Table 3 for application layer modelling of internet traffic.

4.2 Modelling timescales and topology

• In models that capture network at small-time scales, the impact of user behaviour is minimal while the impact of network feedback is dominant. In this case closed-loop is more adequate than open-loop. (Self-similar scaling applies) [24].
• In models that capture large-timescale behaviour: the impact of user is dominant while the impact of network feedback is minimal. In this case open-loop should be adequate.
• At small timescales how the packets are sent into the network within a session matters. (Multi-fractal scaling applies) [24]. Most appropriate models to use are still an open question.
• When modelling elastic traffic (TCP) and investigating packet-scale performance one needs to bear in mind that this is influenced by flow level traffic dynamics, that is performance may deteriorate as the number of flows increases. Whereas flow-scale performance of elastic traffic depends upon traffic demands (bits/sec), exceeding or not available capacity [61].
• A possible approach to model topology is to start with a small simulation (with a few nodes) and then gradually increase the size of the network investigated. These simulations need to be run under somewhat similar conditions. This means that it is necessary to find invariants that are constant as the network size increases. A possible invariant could be the load on links [59].
• If the topology investigated is small and protocol well specified then comparison could be made with real life experiments.

Application		**Distribution**	
	Inter-arrival	***Duration***	***Data***
TELNET	Exponential	Log-normal	Pareto
WWW	Exponential	Log-normal	Self-similar
FTP	Exponential	Log-normal	Pareto
SMTP	Exponential	Log-normal	Log-normal

Table 3: Distributions that model different aspects of internet applications

5 Conclusions

The complexity and heterogeneity of current and future networks requires increasingly sophisticated mathematical and simulation models. There is a growing demand from within the networking research community for common and realistic network models. This is because networking research is likely to benefit from a common platform that allows result comparisons and avoids known pitfalls. Clearly, no single model is universally applicable. In fact, network models should be crafted by the researcher to capture the factors which have greatest impact on the problem at hand. Nonetheless, certain problems are more suitably analysed by means of certain models rather than others. For this reason, researchers would benefit from specific models designed to capture given topological aspects, traffic characteristics, congestion levels and protocol interactions. This paper has provided an introductory (and by no means exhaustive) overview of the current state of the art in the fields of internet traffic and topology modelling. In addition, the paper has reported on existing network models and some proposed new models. Finally a summary has been given of some of the lessons learned in modelling networks and some guidelines for good modelling practice have been given. Ultimately though, the researcher must be aware that the models to use in practice will never be able to reproduce faithfully either the present or future internet. The researcher who aims to gain insight into networks and protocols behaviour should therefore be reconciled to the idea of operating in an environment where a trade-off between simplicity and accuracy needs to be made.

References

[1] A. Adas and A. Makherjee. On resource management and QoS guarantees for long-range dependent traffic. In *Proc. IEEE INFOCOM*, pages 779–787, 1995.

[2] R. Addie, M. Zukerman, and T. Neame. Fractal traffic: measurements, modelling, and performance evaluation. In *Proc. IEEE INFOCOM*, pages 977–984, 1995.

[3] W. Aiello, F. Chung, and L. Lu. A random graph model for massive graphs. In *Proc. of the 32nd Annual Symposium on Theory of Computing*, 2000.

[4] R. Albert and A.-L. Barabasi. Topology of evolving networks: Local events and universality. *Phy. Rev. Letters*, 85(85):5234–5237, 2000.

[5] K. Anagnostakis, M. Greenwald, and R. Ryger. On the sensitivity of network simulation to topology. In *Proc. of the 10th IEEE/ACM Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunications Systems (MASCOTS)*, 2002.

[6] J.-M. Bardet, G. Lang, G. Oppenheim, A. Phillipe, S. Stoev, and M. S. Taqqu. Semi-parametric estimation of the long-range dependence parameter: A survey. In P. Doukhan, G. Oppenheim, and M. S. Taqqu, editors, *Theory and Applications of Long-Range Dependence*, pages 557–577. Birkhäuser, 2003.

[7] M. Barenco and D. K. Arrowsmith. The autocorrelation of double intermittency maps and the simulation of computer packet traffic. *Dynamical Systems*, 19(1):61–74, 2004.

[8] J. Beran. *Statistics for Long-Memory Processes.* Chapman & Hall, NY, 1994.

[9] S. Bertolotti and L. Dunand. OPNET 2.4: an environment for communication network modelling and simulation. In *Proc. of European Simulation Symposium*, 1993.

[10] F. Bonomi and K. Fendick. The rate-based flow control for available bit rate ATM service. *IEEE Networks*, 9(2):24–39, 1995.

[11] M. Borella and G. Brewster. Measurement and analysis of long-range dependent behavior of internet packet delay. In *Proc. IEEE INFOCOM*, pages 497–504, 1998.

[12] R. G. Clegg. *Statistics of Dynamic Networks.* PhD thesis, Dept. of Math., Uni. of York., York., 2004. Available online at: `www.richardclegg.org/pubs/thesis.pdf`.

[13] R. G. Clegg. A practical guide to measuring the Hurst parameter. In N. Thomas, editor, *21st UK Performance Engineering Workshop, School of Computing Science Technical Report Series, CS-TR-916, University of Newcastle*, pages 43–55, 2005. Available online at: `www.richardclegg.org/pubs/rgcpew05.pdf`.

[14] R. G. Clegg and M. M. Dodson. A markov based method for generating long-range dependence. *(To appear) Phys. Rev. E*, 2005.

[15] J. Cowie, D. Nicol, and A. Ogielski. Modelling the global internet. *Computing in Science and Engineering*, 1(1):42–50, Jan-Feb 1999.

[16] M.E. Crovella and A. Bestavros. Explaining world wide web traffic self-similarity. Technical report, TR-95-015 Computer Science Dept., Boston University, 1995.

[17] M.E. Crovella and A. Bestavros. Self-similarity in the WWW traffic: evidence and possible causes. *IEEE/ACM Trans. on Networking*, 5:835–846, 1997.

[18] C. Cunha, A. Bestavros, and M. Crovella. Characteristics of WWW client-based traces. Technical report, TR-95-010, Boston University Computer Science Department, 1995.

[19] M. Doar. A better model for generating test networks. In *Proc. of GLOBECOM*, pages 86–93, 1996.

[20] A. Erramilli, M. Roughan, W. Willinger, and D. Veitch. Self-similar traffic and network dynamics. In *Proc. of the IEEE*, volume 90(5), pages 800–819, 2002.

[21] A. Erramilli, R. P. Singh, and P. Pruthi. An application of deterministic chaotic maps to model packet traffic. *Queueing Systems*, 20:171–206, 1995.

[22] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. In *Proc. ACM SIGCOMM*, pages 251–262, New York, NY, USA, 1999. ACM Press. Available online at: `doi.acm.org/10.1145/316188.316229`.

[23] 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. In *Proc. ACM SIGCOMM*, pages 301–313, 1999.

[24] A. Feldmann, A. C. Gilbert, and W. Willinger. Data networks as cascades: investigating the multifractal nature of internet wan traffic. In *Proc. ACM SIGCOMM*, pages 42–55, 1998.

[25] D. R. Figueiredo, B. Liu, A.Feldmann, V. Misra, D. Towsley, and W. Willinger. On TCP and self-similar traffic. *Elsevier Performance Evaluation Journal, special issue on Long- Range Dependence and Heavy Tail Distributions*, 2005.

[26] D. R. Figueiredo, B. Liu, V. Misra, and D. Towsley. On the autocorrelation structure of TCP traffic. Technical Report 00-55, Department of Computer Science, University of Masssachusetts, 2000.

[27] S. Floyd. Internet research needs better models. Available online at: `www.icir.org/models/bettermodels.html`.

[28] S. Floyd. Connections with multiple congested gateways in packet-switched networks. part 1: One-way traffic. *ACM Comp. Comm. Review*, 21(5):30–47, 1991.

[29] S. Floyd and Van Jacobson. Random early detection gateways for congestion avoidance. *IEEE/ACM Trans. on Networking*, 1(4):397–413, 1993.

[30] S. Floyd and E. Kohler. Internet research needs better models. In *Proc. of HotNets-I*, 2002.

[31] M. Garrett and W. Willinger. Analysis, modelling, and generation of self-similar VBR video traffic. In *Proc. ACM SIGCOMM*, 1994.

[32] M. Grossglauser and J.C. Bolot. On the relevance of long-range dependence in network traffic. In *Proc. ACM SIGCOMM*, pages 15–24, 1996.

[33] J. Heidemann, K. Mills, and S. Kumar. Expanding confidence in network simulations. *IEEE Network Magazine*, 15(5):58–63, 2001.

[34] N. Hohn, D. Veitch, and P. Abry. Does fractal scaling at the IP level depend on TCP flow arrival processes? In *Proc. of The Second Internet Measurement Workshop*, 2002.

[35] B. A. Huberman and R. M. Lukose. Social dilemmas and internet congestion. *Science*, 277:535–537, 1997.

[36] H. Jiang and C. Dovrolis. Passive estimation of TCP round-trip times. *ACM Comp. Comm. Review*, 32(3), 2002.

[37] C. Jin, Q. Chen, and S. Jamin. Inet: Internet topology generator. Technical Report CSE-TR-433-00, EECS Department, University of Michigan, 2000.

[38] S. Keshav. Packet-pair flow control. Available online at: `www.cs.cornell.edu/skeshav/doc/94/2-17.ps`, 1994.

[39] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. On the self-similar nature of Ethernet traffic. In D. P. Sidhu, editor, *Proc. ACM SIGCOMM*, pages 183–193, San Francisco, California, 1993.

[40] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. On the self-similar nature of ethernet traffic (extended version). *IEEE/ACM Trans. on Networking*, 2(1):1–15, 1994.

[41] B. B. Mandelbrot. *The Fractal Geometry of Nature*. W. H. Freeman and Company, 1982.

[42] S. McCann and S. Floyd. The LBNL network simulator, 1997. Available online at:
`http://www.isi.edu/nsnam`.

[43] A. Medina, A. Lakhina, I. Matta, and J. Byers. BRITE: An approach to universal topology generation. In *Proc. of MASCOTS*, 2001.

[44] I. Norros. A storage model with self-similar input. *Queuing Systems*, 16:387–396, 1994.

[45] K. Park. On the effect and control of self-similar network traffic: A simulation perspective. In *Proc. of Winter Simulation Conference (Atlanta, GA)*, pages 989–996, 1997.

[46] K. Park and W. Willinger. *Self-Similar Network Traffic and Performance Evaluation.* Wiley, NY, 2000.

[47] Kihong Park, Gitae Kim, and Mark Crovella. On the relationship between file sizes, transport protocols, and self-similar network traffic. Technical Report 1996-016, Boston University, 1996.

[48] M. Parulekar and A Makowski. Tail probabilities for a multiplexer with self-similar traffic. In *Proc. IEEE INFOCOM*, pages 1452–1459, 1996.

[49] V. Paxson. Empirically derived analytic models of wide-area TCP connections. *IEEE/ACM Trans. on Networking*, 2(4):316–336, 1994.

[50] V. Paxson. Fast, approximate synthesis of fractional gaussian noise for generating self-similar network traffic. *Computer Comm. Rev.*, 27:5–18, 1997.

[51] V. Paxson and S. Floyd. Why we don't know how to simulate the internet. In *Winter Simulation Conference*, pages 1037–1044, 1995.

[52] V. Paxson and S. Floyd. Wide-area traffic: the failure of Poisson modeling. *IEEE/ACM Trans. on Networking*, 3:226–244, 1995.

[53] J. M. Peha. Retransmission mechanisms and self-similar traffic models. In *IEEE/ACM/SCS Comm. Networks and Distributed Systems Modelling and Simulation Conference*, pages 47–52, 1997.

[54] K. Perumalla, R. Fujimoto, and A. Ogielski. TeD a language for modelling telecommunications networks. *ACM SIGMETRICS Performance Evaluation Review*, 25, 1998.

[55] D. Rao and P. Wilsey. Simulation of ultra-large communication networks. In *Proc. Of Seventh International Symposium on Modelling, Analysis, and Simulation of Computer and Telecommunication Systems*, 1999.

[56] R. H. Reidi and J. Lévy Véhel. Multifractal properties of TCP traffic: a numerical study. Technical Report 3129, INRIA, 1997. Available online at
`www.stat.rice.edu/~riedi/Publ/PDF/t.pdf`.

[57] R. H. Riedi. Multifractal processes. In P. Doukhan, G. Oppenheim, and M. S. Taqqu, editors, *Theory And Applications Of Long-Range Dependence*, pages 625–716. Birkhäuser, 2003.

[58] R. H. Riedi, M. S. Crouse, V. J. Ribeiro, and R. G. Baraniuk. A multifractal wavelet model with application to network traffic. *IEEE Special Issue On Information Theory*, 45(April):992–1018, 1999.

[59] G. Riley and M. Ammar. Simulating large networks: How big is big enough? In *Proc. of First International Conference on Grand Challenges for Modeling and Simulation*, 2002.

[60] G. Riley, R. Fujimoto, and M. Ammar Georgia Tech. Parallel/distributed ns, 2000. Available online at:
`www.cc.gatech.edu`.

[61] J. Roberts. Traffic theory and the internet. *IEEE Comm. Magazine*, January 2001.

[62] B. K. Ryu and A. Elwalid. The importance of long-range dependence of VBR video traffic in ATM traffic engineering: Myths and realities. *ACM Comp. Comm. Review*, 26:3–14, 1996.

[63] R. V. Solé and S. Valverde. Information transfer and phase transitions in a model of internet traffic. *Physica A*, 289, 2001.

[64] Tangmunarunkit, R.Govindan, S. Jamin, S. Shenker, and W. Willinger. Network topology generators: degree-based vs. structural. In *Proc. of the 2002 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 147–159, 2002.

[65] M. S. Taqqu and V. Teverovsky. Robustness of Whittle type estimators for time series with long-range dependence. *Stochastic Models*, 13:723–757, 1997.

[66] M. S. Taqqu, W. Willinger, and R. Sherman. Proof of a fundamental result in self-similar traffic modelling. *Computer Comm. Rev.*, 27:5–23, 1997.

[67] B. Tsybakov and N.D. Georganas. Self-similar traffic and upper bounds to buffer overflow in an ATM queue. *Performance Evaluation*, 36:359–386, 1999.

[68] A. Veres and M. Boda. The chaotic nature of TCP congestion control. In *Proc. IEEE INFOCOM*, pages 1715–1723, 2000.

[69] X. J. Wang. Statistical physics of temporal intermittency. *Phys. Rev. A*, 40(11):6647–6661, 1989.

[70] B. M. Waxman. Routing of multipoint connections. *IEEE Journal of Selected Areas in Communication*, 6(9):1617–1622, 1988.

[71] N. Weiner. Generalized harmonic analysis. *Acta. Math.*, 55:178–258, 1930.

[72] W. Willinger, M. Taqqu, R. Sherman, and D. Wilson. Self-similarity through high variability: statistical analysis of Ethernet LAN traffic at the source level. *IEEE/ACM Transactions on Networking*, 5(1):71–86, 1997.

[73] N. Wisitpongphan and J. M. Peha. Effect of TCP on self-similarity of network traffic. In *Proc. of 12th IEEE International Conference on Computer Communications and Networks*, 2003.

[74] M. Woolf, D. K. Arrowsmith, R. J. Mondragon, and J. M. Pitts. Optimization and phase transitions in a chaotic model of data traffic. *Phys. Rev. E*, 66(4), 2002.

[75] M. Yuksel, B. Sikdar, K. Vastola, and B. Szymanski. Workload generation for ns simulations of wide area networks and the internet. In *Proc. of Communication Networks and Distributed Systems Modeling and Simulation Conference*, 2000.

[76] E. Zegura, Thoughts on Router-level Topology Modeling. The End-to-end interest mailing list, January 2001. Available online at `www.postel.org/pipermail/end2end-interest/2001-January/000033.html`.

[77] X. Zeng, R.Bagrodia, and M. Gerla. GloMoSim: a library for parallel simulation of large-scale wireless networks. In *Proc. of 12th Workshop on Parallel and Distributed Simulations*, 1998.