Local Area Networks and Self-similar Traffic

Hadi Larijani

Dept. of Communication, Network & Elec. Eng.
School of Eng. and Computing
Glasgow Caledonian University, UK
hla@gcal.ac.uk

Abstract. Ethernet is one of the most popular LAN technologies. The capacities of Ethernet have steadily increased to Gbps and it is also being studied for MAN implementation. With the discovery that real network traffic is self-similar and long-range dependent, new models are needed for performance evaluation of these networks. One of the most important methods of modelling self-similar traffic is Pseudo self-similar processes. The foundations are based on the theory of decomposability, which was developed approximately 20 years ago. Many researchers have revisited this theory recently and it is one of the building blocks for self-similar models derived from short-range dependent processes. In this paper we will review LANs, self-similarity, several modelling methods applied to LAN modelling, and focus on pseudo self-similar models.

Keywords: Ethernet, self-similarity, decomposability, pseudo self-similar processes.

1 Introduction

Performance modeling of computer networks is essential to predict the effects of increases in traffic and to allow network managers to plan the size of upgrades to equipment. Historically the assumption has been made that traffic followed Poisson assumptions. That is, the probability of a packet arriving for onward transmission during any short interval is independent of the arrivals in any of the intervals, and depends only on the mean arrival rate and the length of the interval considered. Similarly, the length of packets is usually considered to be exponentially distributed. Both these assumptions are false, but models using them have been successfully validated, mostly in the wide area network field.

Recent measurements of network teletraffic have revealed properties which may have significant consequences to the modelling of computer networks, especially Broadband ISDN and ATM networks. Although self-similarity is not a new concept to the teletraffic community and its origins date back to Mandelbrot's paper in 1965 [1], its consequences were not fully appreciated till the 1990's. One of the most highly acclaimed papers which might be considered as the spark to an explosion of research into this area and its wide scope of applications is the fascinating paper written by Leland, Willinger, Taqqu and Wilson [2]. As Stallings [3] very well stated, this paper rocked the field of network performance modelling and it is arguably the most

D. Kouvatsos (Ed.): Next Generation Internet, LNCS 5233, pp. 174–190, 2011.

important networking paper of the decade. The main finding of this paper is that real traffic does not obey the Poisson assumptions that have been used for years for analytical modelling. ‘Real’ network traffic is more bursty and exhibits greater variability than previously suspected. The paper reported the results of a massive study of Ethernet traffic and demonstrated that it has self-similar statistical properties at a range of time scales: milliseconds, seconds, minutes, hours, even days and weeks. What this means is that the network looks the same when measured over time intervals ranging from milliseconds to minutes to hours. It was later found that self-similarity also occurs in ATM traffic, compressed digital video streams, World Wide Web (WWW) traffic, Wide Area Networks (WAN) traffic, etc., [4]. Self – similar traffic is very different from both conventional telephone traffic and from the currently accepted norm for models of packet traffic. Conclusions of recent empirical studies regarding the nature of network traffic have all concurred on one issue: the data is self – similar in nature. What are the ramifications of this discovery? Consequently, the following needs to be addressed:

1. What are the performance implications of self – similar data traffic upon telecommunication systems?
2. How can researchers utilize queuing models to study this behavior?

Note that some researchers do not believe that queuing models are sufficient and suggest research on new tools for this end [5]. Self – similarity has immense impact on a wide variety of fields such as: traffic modelling, source characterization, performance evaluation, analytical modelling, buffer sizing, control mechanisms, etc. For example, Partridge [6] foresaw the implications of this in congestion control and stated ‘anyone interested in congestion control should read the paper’ [2]. A.A. Kherani [7] has looked into the effect of adaptive window control in LRD network traffic. His study indicates that the buffer behaviour in the Internet may not be as poor as predicted from an open loop analysis of a queue fed with LRD traffic; and it shows that the buffer behaviour (and hence the throughput performance for finite buffers) is sensitive to the distribution of file sizes.

A. Veres et al. [8] analyzed how TCP congestion control can propagate self-similarity between distant areas of the Internet. This property of TCP is due to its congestion control algorithm, which adapts to self-similar fluctuations on several timescales. The mechanisms and limitations of this propagation were investigated. It was demonstrated that if a TCP connection shared a bottleneck link with a self-similar background traffic flow, it propagates the correlation structure of the background traffic flow asymptotically, above a characteristic timescale. The cut-off timescale depends on the end-to-end path properties, e.g. round-trip time and average window size, and the receiver window size in case of high-speed connections. It was also shown that even short TCP connections can propagate long-range correlations effectively. In case when TCP encounters several bottleneck hops, the end-user perceived end-to-end traffic was also long-range dependent and it was characterized by the largest Hurst exponent. Through simple examples, it was shown that self-similarity of one TCP stream can be passed on to other TCP streams that it was multiplexed with.

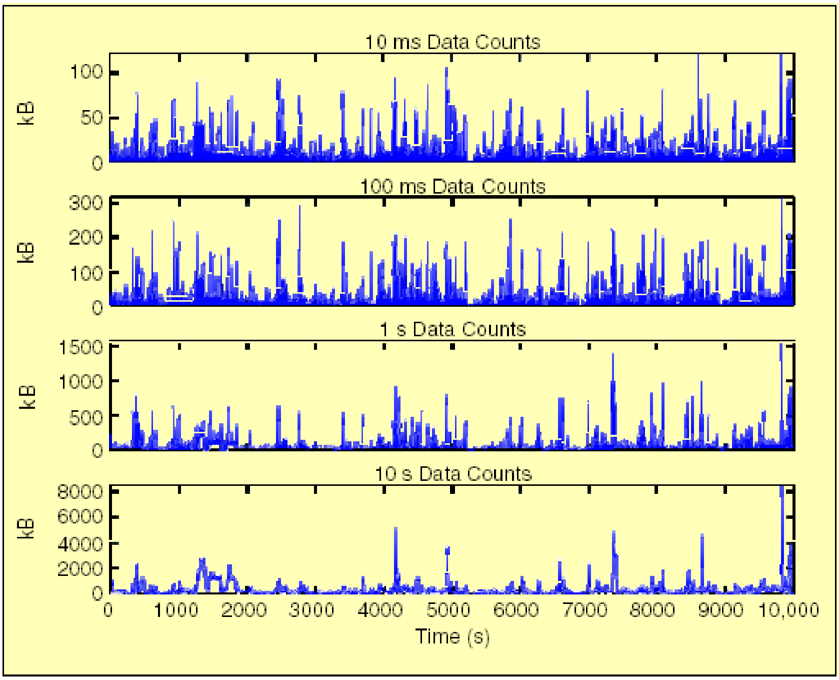
1.1 What Is Self-similarity?

Self-similarity and fractals are notions pioneered by B.B. Mandelbrot [9]. He describes the phenomenon where a certain property of an object – for example, a natural image, the convergent subdomain of certain dynamical systems, a time series (the mathematical object of interest) – is preserved with respect to scaling in space and time. If an object is self-similar or fractal, its parts, when magnified, resemble – in a suitable sense – the shape of the whole object.

Stochastic self-similarity is of more importance to our study of network traffic. Analysis and modelling of. computer network traffic is a daunting task considering the amount of available data. This is quite obvious when considering the spatial dimension of the problem, since the number of interacting computers, gateways and switches can easily reach several thousands, even in a local area network (LAN) setting. This is also true on the time dimension: Willinger and Paxson in [10] cite the figures of 439 million packets and 89 gigabytes of data for a single week record of the activity of a university gateway in 1995. The complexity of the problem further increases when considering wide area network (WAN) data [11]. In light of the above, it is clear that a notion of importance for modern network engineering is that of invariants, i.e., characteristics that are observed with some reproducibility and independently of the precise settings of the network under consideration. In this study we focus on one such invariant related to the time dimension of the problem, namely, long-range dependence or self-similarity. A striking feature, which collaborate the conjecture that self-similarity, long-range dependence, and heavy-tailness are really meaningful traffic invariants, is that they can be observed, to some extent, without using any specific experimental protocol. Accordingly, the traffic data in Figure 1 corresponds to actual 100 Mb/s Ethernet traffic, which was measured on a server in Drexel University [12]. To generate this trace, all packets of private connections with this server, broadcasting, and multicasting were captured and time-stamped during several hours. Cappe et. al. [12] only consider byte counts (size of the transferred data) measured on 10 ms intervals, which is the data represented in the top plot of Figure 1. The overall length of the record is about three hours (exactly, 10^4 s). The three other plots in Figure 1 correspond to the "aggregated" data obtained by accumulating the data counts on increasing time intervals. The striking feature in it is that the aggregation is not really successful in smoothing out the data. The aggregated traffic still appears bursty in the bottom plot despite the fact that each point in it is obtained as the sum of one thousand successful values of the series displayed in the top plot of Figure 1.

Similar characteristics have been observed in many different experimental setups, including both LAN and WAN data (e.g., [13], [14], [15], and the references therein).

Unlike deterministic fractals, the objects corresponding to Figure 1 do not possess exact resemblance of their parts with the whole at finer details. Here, we assume that the measure of "resemblance" is the shape of a graph with the magnitude suitably normalized. Indeed, for measured traffic traces, it would be too much to expect to observe exact, deterministic self-similarity given the stochastic nature of many network events (e.g., source arrival behaviour) that collectively influence actual network traffic. If we adopt the view that traffic series are sample paths of stochastic processes and relax the measure of resemblance, say, by focusing on certain statistics of the rescaled time series, then it may be possible to expect exact self-similarity of the mathematical objects and approximate similarity of their specific realizations with respect to these relaxed measures. Second - order statistics are statistical properties that capture burstiness or variability, and the autocorrelation function is a yardstick with respect to which scale invariance can be fruitfully defined. The shape of the autocorrelation function – above and beyond its preservation across rescaled time series – will play an important role. In particular, correlation, as a function of time lag, is assumed to decrease polynomially as opposed to exponentially. The existence on nontrivial correlation "at a distance" is referred to as *long-range dependence*.



▲ *1. The Drexel data (10,000 s in total) viewed through four different aggregation intervals: from top to bottom, 10 ms, 100 ms, 1 s, and 10 s.*

Fig. 1. Example of Self-Similar Traffic Trace

1.2 Self-similar Processes: Basic Definitions

A self-similar process is invariant in distribution under scaling of time. Intuitively, if we look at several pictures of a self-similar process at different time scales they will all look *similar*. Figure 1 shows the visual differences between distributions of traditional models and self-similar processes in a few scales. There are a number of different, non-equivalent definitions of self-similarity. The standard one states that a continuous time process $Y = \{Y(t), t \in T\}$ is *self-similar* (with self-similarity parameter *H*) if it satisfies the condition:

$$Y(t) \underset{=}{d} \alpha^{-H} Y(\alpha t), t \in T, \forall \alpha > 0, 0 \leq H < 1 \qquad (1)$$

where the equality is in the sense of finite-dimensional distributions. *H* is known as the *Hurst* parameter, in honour of an early pioneer of the study of self – similarity [16]. While a process *Y* satisfying (1) can never be stationary, it is typically assumed to have stationary increments.

A second definition of self-similarity, more appropriate in the context of standard time series theory, involves a stationary sequence $X = \{ X(t), i \geq 1 \}$.

Let

$$X^{(m)}(k) = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X(i) \quad k = 1, 2, \ldots, \qquad (2)$$

be the corresponding aggregated sequence with a level of aggregation *m*, obtained by dividing the original series *X*, into non-overlapping blocks of size *m* and averaging every block. The index, *k*, labels the block. If *X* is the increment process of a self-similar process *Y* defined in (1) i.e., $(X(i) = Y(i+1) - Y(i))$, then for all integers *m*,

$$X \underset{=}{d} m^{1-H} X^{(m)} \qquad (3)$$

A stationary sequence $X = \{X(i), i \geq 1\}$ is called *exactly self- similar* if it satisfies (3) for all aggregation levels *m*. The second definition of self-similarity is closely related (but not equivalent) to the first. A stationary sequence $X=\{X(i), i \geq 1\}$ is said to be *asymptotically self-similar* if (3) holds as $m \to \infty$. Similarly, we call a covariance-stationary sequence $X = \{X(i), i \geq 1\}$ *exactly second-order self-similar* or *asymptotically second-order self-similar* if $m^{1-H}X^{(m)}$ has the same variance and autocorrelation as *X*, for all *m*, or as $m \to \infty$.

Self-similarity is often investigated not through the equality of finite-dimensional distributions, but through the behaviour of the absolute moments. Thus, a third definition of self-similarity (implied by but not equivalent to the second definition) is simply that the moments must scale. Thus consider:

$$\mu^{(m)}(q) := E \mid X^{(m)} \mid^q = \left| \frac{1}{m} \sum_{i=1}^{m} X(i) \right|^q \qquad (4)$$

If *X* is self-similar, then $\mu^{(m)}(q)$ is proportional to $m^{\beta(q)}$, i.e., $\log \mu^{(m)}(q)$ is linear to $\log m$.

For a fixed *q:*

$$\log \mu^{(m)}(q) = \beta(q) \log m + C(q) \qquad (5)$$

In addition, the exponent $\beta(q)$ is linear with respect to *q*. In fact, since $X^{(m)}(i) \underset{=}{d} m^{H-1} X(i)$, we have: $\beta(q) = q(H-1)$.

1.3 Long-Range Dependence

A stochastic process satisfying the following relation is said to exhibit *long-range dependence* (see [17] or [18]):

Let $X = (X_t : t = 0,1,2,\ldots)$ be a *covariance stationary* (sometimes called *wide-sense stationary*) stochastic process with mean μ, variance σ^2 and auto correlation function

$$r(\kappa) \sim \kappa^{-\beta} L_1(\kappa) \qquad \text{as } \kappa \rightarrow \infty, \tag{6}$$

where $0 < \beta < 1$ and L_1 is slowly varying at infinity, that is $\lim_{t\rightarrow\infty} L_{1(tx)/} L_{1(x)} = 1$ for all x>0.

1.4 Ethernet

Ethernet is the most widely used local area (LAN) technology. It was used to fill the middle ground between long – distance, low – speed networks and specialized, computer – room networks carrying data at high speeds for very limited distances. Ethernet is well suited to applications where a local communication medium must carry sporadic, occasionally heavy traffic at high peak data rates.

Ethernet network architecture has its origins in the 1960's at the University of Hawaii, where the earliest and simplest random access scheme was developed, (pure-ALOHA) [19]. Another more efficient random access scheme called CSMA (carrier sense multiple access) was developed by Kleinrock's [20] team at UCLA. Ethernet uses an access method called carrier sense multiple access/ collision detection (CSMA/ CD), which was developed at Xerox corporation's Palo Alto Research Center (PARC) in the early 1970's. This was used as the basis for the Institute of Electrical and Electronic Engineers (IEEE) 802.3 specification released in 1980.Shortly after the 1980 IEEE 802.3 specification, Digital Equipment Corporation, Intel Corporation, and Xerox Corporation jointly developed and released Ethernet specification version 2.0, that was substantially compatible with IEEE 802.3.

Together, Ethernet and IEEE 802.3 currently maintain the greatest market share of any LAN protocol. Today, the term Ethernet is often used to refer to all carrier sense multiple access/ collision detection (CSMA/CD) LAN's that generally conform to Ethernet specifications, IEEE 802.3.

1.5 Self-similar Traffic Modeling

It has been known for a long time that network traffic is self – similar in nature. In fact, Mandelbrot was the first to apply the self- similarity concept to the analysis of communication systems [1]. As a consequence of Leland's et al. paper [13] much work has recently appeared addressing various aspects of self-similarity [21]. This research can be classified into the following three categories:

<u>Network Traffic Trace Analysis:</u> This research typically analyses traffic traces from production networks so that statistical tools can be employed to identify the presence of self – similarity. This genre of papers indicates that Long Range Dependence (LRD) is an omnipresent phenomenon encompassing both local area and wide area network

traffic. Furthermore, work in this area has consistently demonstrated that sources such as the WWW and VBR video services exhibit self – similar properties [21].

Simulation and Analytical Models: Research in this area attempts to investigate the effect of self – similar data traffic upon telecommunication systems using either simulation or asymptotic analytical models. These papers conclude that data traffic with properties of self – similarity and LRD significantly degrade system performance. One important result in this area has been the development of Fractional Brownian Noise (FBN) models. FBN models not only capture properties of self – similarity but LRD in the counting process. Consequently, many researchers have utilized computer simulations of FBN models to study the impact of LRD upon queuing behaviour. Typical analytical approaches attempt to find asymptotic bounds concerning selected performance characteristics of the queue (e.g. buffer overflow probability) [22].

Conceptual Analysis: The third category of papers attempts to physically understand how self - similarity arises in production networks. One model proposed by Willinger, Taqqu, and Sherman which attempts to address this issue utilizes an ON/OFF source model (also commonly known as “packet train model”) [23]. Their ON/OFF model purports that self – similarity arises as a consequence of independent contributions from power – tail sources. The reason is that mathematical analysis indicates that the superposition of many power – tail ON/OFF sources with alternating ON and OFF periods produces aggregate network traffic which exhibits properties of self-similarity and LRD. The authors conclude that their ON/OFF model is successful in describing characteristics of the measured traffic.

Various methods have been presented for modelling self-similar processes. The two major families of self-similar time series models are fractional Gaussian noise (i.e. increment processes of fractional Brownian motion) and fractional ARIMA processes (auto-regressive integrated moving-average), (a generation of the popular ARIMA time series models). Other stochastic approaches to modelling self-similar features that have been presented are:

- Shot-noise processes
- Linear models with long-range dependence
- Renewal reward processes and their superposition
- Renewal processes or ``zero-rate'' processes
- Aggregation of simple short-range-dependent models
- Wavelet analysis
- Approaches based on the theory of chaos and fractals
- Batch Renewal Process

Willinger et al have provided an excellent comprehensive review in [24]. R. Fretwell and D. Kouvatsos [25] use the batch renewal process for both LRD and SRD traffic. They also show some applications of the batch renewal process in simple queues and in queuing network models.

Generally we can group the different approaches researchers take in modeling self-similarity into two distinct ‘camps’:

- One group tries to develop approaches that attempt mimicking LRD with the help of short-range dependent models (e.g. [26]).
- The second group tries to advent a new set of tools for modeling self-similar processes (e.g. [27]).

K. Maulik and S. Resnick [28] attempt to model this phenomenon, by a model that connects the small time scale behavior with behavior observed at large time scales of bigger than a few hundred milliseconds. There have been separate analyses of models for high speed data transmissions, which show that appropriate approximations to large time scale behavior of cumulative traffic are either fractional Brownian motion or stable Lévy motion, depending on the input rates assumed. Their paper tries to bridge this gap and develops and analyzes a model offering an explanation of both the small and large time scale behavior of a network traffic model based on the infinite source Poisson model. Previous studies of this model have usually assumed that transmission rates are constant and deterministic.

They considered a nonconstant, multifractal, random transmission rate at the user level which results in cumulative traffic exhibiting multifractal behaviour on small time scales and self-similar behaviour on large time scales.

We follow the first approach, and a Markov chain model which shows self-similarity is developed, based on ideas presented by Robert and LeBoudec [29].

There are two levels of modeling: application and aggregate level. Although it is true that network traffic is governed by many physical factors a ‘good’ model should incorporate those features which are relevant for the problem under consideration. Some of the many factors affecting network traffic flows are:

- user behavior
- data generation, organisation, and retrieval
- traffic aggregation
- network controls
- network evolution

In the Section 2 Pseudo Self-Similar models will be discussed, and a numerical solution suggested for the queuing behavior of a Self-Similar LAN. Experiments on a live Ethernet network will be presented in Section 3 with validation by Opnet simulation application. Finally conclusions will be presented in Section 4.

2 Pseudo Self-similar Models

2.1 Foundations of the Model

Courtois’s [30] theory of decomposability is based on the important observation that large computing systems can effectively be regarded as nearly decomposable systems. Systems are arranged in a hierarchy of components and subcomponents with strong interactions within components at the same level and weaker interactions between other components. Near decomposability has been observed in domains such as: in economics, biology, genetics and social sciences. The pioneers of this theory are

Simon and Ando [31]. What they stated is that aggregation of variables in a nearly decomposable system; we must separate the analysis of the short - term and the long - term dynamics. They proved two major theorems. The first says that a nearly decomposable system can be analyzed by a completely decomposable system if the intergroup dependences are sufficiently weak compared to the intragroup ones. The second theorem says that even in the long - term, the results obtained in the short - term will remain approximately valid in the long - term, as far as the relative behavior of the variables of the same group is concerned. Robert and Le Boudec [29] state that LAN traffic is composed of different timescales. The Markov chain proposed is in fact decomposable at several levels. In a first step, the development is done for only one level of decomposability. The Markov chain to consider is presented in section 2.3 and it is characterized by its transition matrix $(n*n)A$ and its state probabilities, $\pi(\pi_{t+1} = \pi_t A)$, A is nearly decomposable. Let A^* be completely decomposable, then A^* is composed of squared submatrices placed on the diagonal:

$$A^* = \begin{pmatrix} A_1^* & \cdots & 0 & 0 \\ 0 & A_2^* & \cdots & 0 \\ 0 & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & A_N^* \end{pmatrix}$$

The remaining elements are zero. If we apply the first theorem of Simon and Ando [31] we can develop the general form of matrices that are nearly completely decomposable with the form described in section 2.3. A is defined in section 2.3 and $A*$ is defined as below:

$$A^* = \begin{Bmatrix} 1-1/a-1/a^2-\cdots-1/a^{n-1} & 1/a & 1/a^2 & \cdots & 0 \\ q/a & 1-q/a & 0 & \cdots & 0 \\ (q/a)^2 & 0 & 1-(q/a)^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 \end{Bmatrix}$$

with $q\langle a$, A^* is a non-ergodic matrix.

2.2 Pseudo Long-Range Dependent Process

Mathematically, the difference between short – range and long – range dependencies is clear, for a short - range dependent process:

$\sum_{\tau=0}^{\infty} Cov(X_t, X_{t+\tau})$ is convergent

Spectrum at 0 is finite

$Var(X^{(m)})$ is for large m asymptotically of the form $VarX / m$

The averaged process $X_k^{(m)}$ tends to second-order pure noise as $m \to \infty$

For a long-range dependent process:

$$\sum_{\tau=0}^{\infty} Cov(X_t, X_{t+\tau}) \text{ is divergent}$$

Spectrum at 0 is singular

$Var(X^{(m)})$ is of the form $m^{-\beta}$ (for large m asymptotically)

The averaged process $X_k^{(m)}$ does not tend to second-order pure noise as $m \to \infty$.

All stationary autoregressive-moving average processes of finite order an all Markov chains (including semi – Markov processes) are included in the first category. In the second category, we have the fractional Brownian motion, ARIMA processes, and chaotic maps which have long – range dependencies. If we look more closely to these definitions we see that a process having "long – term dependences", but which is limited, is considered as a short – term dependent process. This is exactly the case with Ethernet measurements at Bellcore. If we consider the number of Ethernet packets arriving at a time interval 1 s to be our process, then over 4-5 orders of magnitude we observe long-term dependences. So our process looks the same (distribution wise) for 10, 100, 1000, 10000 s. However, in the order of days, researchers at Bellcore have observed a stabilization of the index of dispersion indicating a lack of self-similarity. So according to our previous definitions, a short - term dependent process would be sufficient to model LAN traffic. The difference with the other processes (Poisson, ON-OFF, etc.) is striking and that is why they should be categorized differently.

Therefore Robert and Le Boudec proposed to name them Pseudo long – range dependent processes:

"A pseudo long – range dependent process is able to model (as well as an (exactly) long – range process) aggregated traffic over several timescales".

This reflects the fact that in practice, we have always a finite set of data, and asymptotic conditions are never met.

2.3 Suggested Process

A discrete time Markov modulated model for representing self-similar data traffic was proposed by Robert and Le Boudec [29]. The cell arrivals on a slotted link is considered: call X_t the random variable representing the number of cells (assumed to be *0* or *1*) during the *t*th time slot, namely during time interval $[t - 1, t)$. Let $Y_t = i$ be the modulator's state i, $i \in 1, 2, 3, \ldots, n$, at time t. The arrivals of the cells are modulated by a n-stated discrete time Markov chain with transition probabilities $\alpha_{ij}(t_1, t_2) = \Pr(Y_{t2} = j \mid Y_{t1} = i)$. Let ϕ_{ij} denote the probability of having j cells in one time slot, given that the modulator's state is i; more specifically $\phi_{ij} = \Pr(Y_t = j \mid Y_t = i)$.

The modulated chain state probabilities are noted as $\pi_{it} = \Pr(Y_t = i)$, i is referred to the modulator's stated and *t* to the time. The Markov modulated chain is assumed stationary and homogeneous.

The Markov chains that they suggested using is the following:

$$A = \begin{Bmatrix} 1-1/a-1/a^2-\cdots-1/a^{n-1} & 1/a & 1/a^2 & \cdots & 1/a^{n-1} \\ q/a & 1-q/a & 0 & \cdots & 0 \\ (q/a)^2 & 0 & 1-(q/a)^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ (q/a)^{n-1} & 0 & 0 & \cdots & 1-(q/a)^{n-1} \end{Bmatrix}$$

$$\Lambda = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

So, the Markov chain has only three parameters: a, q, plus the number of states in the Markov chain *n*.

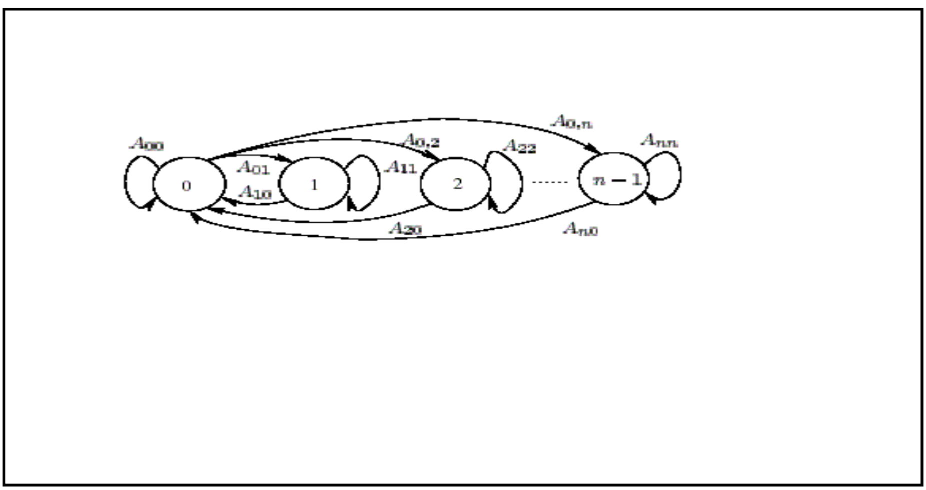


Fig. 2. The state-transition diagram of the modulating Markov chain of the Pseudo Self-similar Traffic (PSST) model

Notice that the parameters α and q need to fulfill certain conditions so that A is indeed a stochastic matrix describing a discrete-time Markov chain; q, α >0, q < α such that $0 \leq A_{0,0} \leq 1$.

In sequel, we denote with $A^k_{i,j}$ the entry in row i and column j of A^k. We furthermore define $N = (N_t, t \in \mathbb{N})$ as the discrete-time stochastic process describing the number of arrivals over time, as described by the Markov-Modulated Bernoulli process (MMBP).

3 Experiments

3.1 Overview of Experiments

As shown in the Figure 3 diagram traces of packets are used as input to a S-Plus script which will evaluate the Hurst parameter using the aggregate variance method. Then another S-Plus script will evaluate parameters: n, q, and *a* in Robert's Pseudo-Self

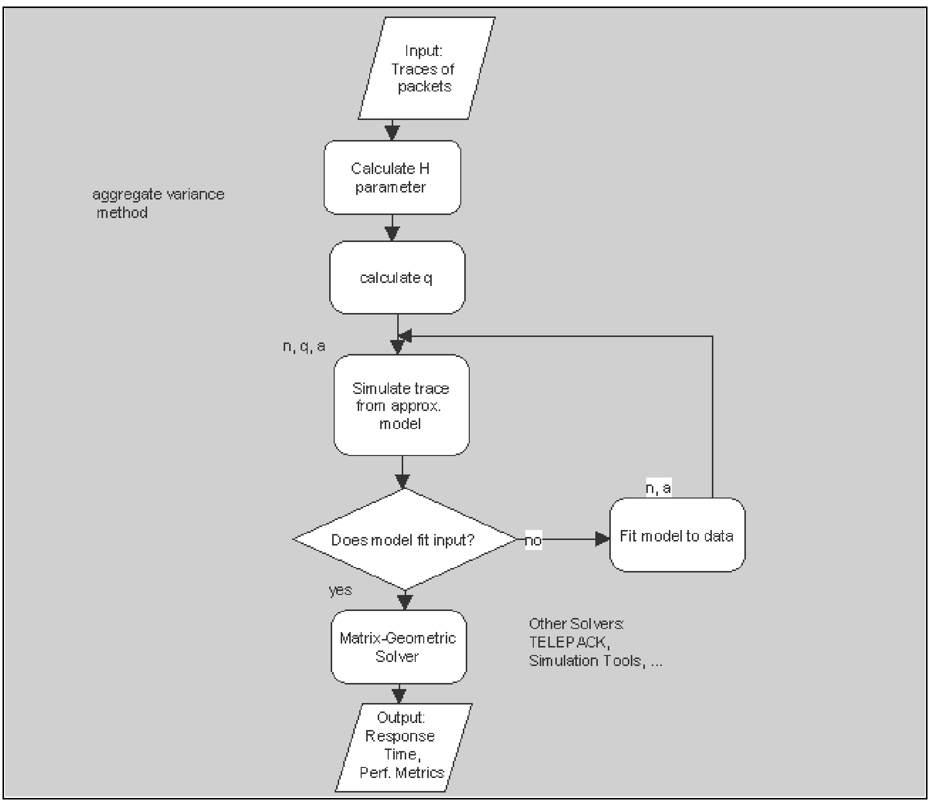


Fig. 3. Overview of Experiments

similar Markov Chain described in chapter 2.3. A close fit to the calculated Hurst parameter from the traffic traces is evaluated iteratively. Then we used the Matrix-Geometric method to calculate the mean queue length and response time.

3.2 Packet Traces

Packet traces were generated by a custom made self-similar packet generator. Traces using Vern Paxson's [32] fast approximation algorithm proved more suitable. Traces were generated with *Hurst parameter* varying from 0.55 to 0.95. Traces sufficient for an hour simulation were prepared. Two different metrics were considered:

- Self-similar packet length
- Self-similar inter-arrival times

3.3 Experiment Results

In the first results we obtained from the experiment and which were published in [33], we showed the effect of different parameters related to Stephen Roberts pseudo self-similar models (in section 2) to mean queue length. More specifically we showed that increases in the a parameter (see section 2.3) consequently increase the mean queue length. This is in agreement with many studies that have proven self-similarity traffic increases mean queue length. Note that increased values of the a parameter translate to higher *Hurst* parameters. The results were obtained by solving a single user queue with arrivals from the pseudo self –similar process with varying *Hurst* parameters.

The obvious conclusion from our study was that modelling queues with self-similar traffic, using traditional Poison arrivals greatly underestimates mean queue length.

In the next two experimental studies [34] and [35] we studied the effect of artificially generating self-similar traffic in an Ethernet LAN on response time. The studies were conducted in an Ethernet where a custom-made packet generator generated packets with specified Hurst parameters. The response time was monitored by a Ping function using a high-resolution clock based on the processor frequency. Results show that response time is in most cases self-similar when the arrival process has a Hurst parameter in the interval (0.55 0.95). In another simulation study using Opnet we validated the three previous experimental studies. Packet length was based on a bi-modal with probabilities based on distributions in [36].

3.4 Novel Numerical Solution Queuing Model

Using the input pseudo self-similar process (PSST) (section 2.3) in a queuing model is possible in a number of ways. First, a discrete time queue could be constructed, and the process used as the input to that queue. Since at most 1 customer arrives per slot, and one customer can be served, this would not be very interesting. A more interesting approach would be to use several of the self similar processes as input, so that queues might have a chance to build up. This would involve construction of a Markov chain with each state representing the states of the individual self similar processes. Construction of such a chain is straightforward, and then the analysis of a discrete time queue with that input process could be conducted.

A second approach, which we followed, is to (incorrectly) assume that the process is continuous, and to assume that when the input process is in state 1 arrivals form a Poisson process with rate 1, and that the server gives exponentially distributed service times with mean 1 whatever the state of the input. This means that we are analyzing an M/M/1 queue in a Markovian environment, which has been the subject of many studies. We solve this queuing model using the matrix-geometric technique. More details can be found in [33].

The state of the system can be denoted by a pair of integers, (I; J), with I representing the number of customers in the system, and J, $1 \leq J \leq n$, representing the state of the input process. When the input process is in state 1, jobs arrive in a Poisson stream at rate 1, and in all other states of the input process there are no arrivals. Whatever the state of the input process, service takes place at rate 1.

The steady state probabilities can be denoted as $\pi_{ij} = \Pr(I = i \wedge J = j)$and can be related using the balance equations, and if the vector $\pi_i = (\pi_{i1,} \pi_{i2,} \pi_{i3,} \ldots,.\pi_{in)}$ is defined, then the balance equations can be expressed as:

$$A_2\pi_{i+1} + A_1\pi_I + A_0\pi_{i-1} \tag{7}$$

The matrices A_2 and A_0 are diagonal matrices, with entries consisting of the arrival and service rates, respectively, in the corresponding state of the input process.

A_1 is $P - I - A_2 - A_0$, where P is where P is the transition matrix of the input process.

Neuts [37] shows that: $\pi_i = R^i \pi_0$.

where R is the unique solution of:

$$A_2R^2 + A_1R + A_0 = 0 \tag{8}$$

And

$$\pi_0 = (1\text{-R})\alpha \tag{9}$$

where α is the steady state distribution of the Markov chain representing the input process.

4 Conclusions

4.1 Effect of Packet Size and Hurst Parameter

In [34] and [35] we showed results of two major experiments based on a novel packet generator and high resolution ping function. The fundamental findings were that response time in experiments with random packet length proved to be self-similar, and experiments with bi-modal proved that response time was non-self-similar. Simulations on similar network with the same traces as the experiments showed that the Hurst parameter has a declining effect on the delay, and that the bi-modal packet length has a similar effect.

4.2 Comparison of Traffic Generator (Measurement Approach) and Simulation Model

The difference in the results of the measurement study and simulation study can be attributed to the following factors:

- Measurement approach was conducted in an open environment with traffic from other sources than the self-similar traffic generator
- Ping function also contributed to traffic, this had a high overhead in lower time resolutions.
- Number of traffic sources were not identical, even though packet traces were identical
- Opnet simulation Kernel had a different effect than the self-similar generator and measurement approach

In summary the measurement study showed the effect of self-similarity can under conditions be passed down to delay, and the simulation study showed how the bimodal packet distribution can affect the delay, this factor is more than the *Hurst* parameter.

References

[1] Mandelbrot, B.B.: Self-similar error clusters in communication systems and the concept of conditional stationarity. IEEE Transactions on Communications Technology Com 13, 71–90 (1965)
[2] Leland, W.E., Taqqu, M.S., Willinger, W., Wilson, D.V.: On the self-similar nature of Ethernet traffic. IEEE/ACM Transactions in Networking 2, 1–15 (1994)
[3] Stallings, W.: Viewpoint: Self-similarity upsets data traffic assumptions. IEEE Spectrum 34, 28–29 (1997)
[4] Willinger, W., Taqqu, M.S., Erramili, A.: A bibliographical guide to self-similar traffic and performance modeling for modern high-speed networks. In: Stochastic Networks: Theory and Applications, pp. 339–366 (1996)
[5] Erramilli, A., Narayan, O., Willinger, W.: Experimental Queueing Analysis with Long-Range Dependent Packet Traffic. IEEE/ACM Transactions in Networking 4, 209–223 (1996)
[6] Partridge, C.: The end of simple traffic models. IEEE Network 7, 3 (1993)
[7] Kharani, A.A., Kumar, A.: Long range dependence in network traffic and the closed loop behaviour of buffers under adaptive window control. Performance Evaluation 61, 95–127 (2005)
[8] Veresa, A., Kenesic, Z., Molnarc, S., Vattayd, G.: TCP's role in the propagation of self-similarity in the Internet. Computer Communications 26, 899–913 (2003)
[9] Mandelbrot, B.B.: The Fractal Geometry of Nature. W.H. Freeman, New York (1982)
[10] Willinger, W., Paxson, V.: Discussion of the paper by S.I. Resnick. Annals in Statistics 25(5), 1856–1866 (1997)
[11] Liu, Z., Niclausse, N., Jalpa-Villanueva, C., Barbier, S.: Traffic model and performance evaluation of web servers. INRIA, Sophia Antipolis, France, Tech. Rep. RR-3840 (1999)
[12] Cappe, O., Moulines, E., Pesquet, J.-C., Petropulu, A., Yang, X.: Long-Range Dependence and Heavy-Tail Modeling for Teletraffic Data. IEEE Signal Processing Magazine 19, 14–27 (2002)

[13] Leland, W.E., Taqqu, M.S., Willinger, W., Wilson, D.V.: On the self-similar nature of Ethernet traffic. Computer Communication Review 23, 183–193 (1993)
[14] Paxson, V., Floyd, S.: Wide-area traffic: The failure of Poisson modeling. IEEE/ACM Trans. Networking 3, 226–244 (1995)
[15] Beran, J., Sherman, R., Taquu, M.S., Willinger, W.: Long-range dependence in variable-bit-rate video traffic. IEEE Trans. Commun. 43, 1566–1579 (1995)
[16] Hurst, H.E.: Long-term storage capacity of reservoirs. Transactions of the American Society of Civil Engineers 116, 770–799 (1951)
[17] Beran, J.: Statistical methods for data with long-range dependence. Statistical Science 7, 404–427 (1992)
[18] Cox, D.R.: Long-range dependence: a review. Statistics: An Appraisal. In: David, H.A., David, H.T. (eds.) pp. 55–74. Iowa State Univ. Press (1984)
[19] Abramson, N.: The ALOHA System- Another alternative for computer. In: Fall Joint Comput. Conf., pp. 281–285. AFIPS Press, Monvale (1970)
[20] Kleinrock, L., Tobagi, F.A.: Packet switching in radio channels: Part 1-Carrier Sense multiple-access modes and their throughput- delay characteristics. IEEE Trans. Commun. Com. 23, 1400–1416 (1975)
[21] Crovella, M.E., Bestavros, A.: Self - similarity in world wide web traffic:evidence and possible causes. IEEE/ACM Transactions on Networking 5(6), 835–846 (1997)
[22] Likhanov, N., Tsybokov, B., Georganas, N.D.: Analysis of an ATM buffer with self - similar (fractal) input traffic. In: Proc IEEE INFOCOM 1995, Boston, pp. 985–992 (1995)
[23] Willinger, W., Taqqu, M.S., Erramili, A.: Self-Similarity in High-Speed Packet Traffic: Analysis and Modeling of Ethernet Traffic Measurements. Statistical Sciences 10, 67–85 (1995)
[24] Willinger, W., Taqqu, M.S., Erramili, A.: A bibliographical guide to self- similar traffic and performance modeling for modern high-speed networks. In: Kelly, F.P., Zacharay, S., Zeidin, I. (eds.) Stochastic Networks: Theory and Applications, pp. 339–366. Oxford University Press, Oxford (1996)
[25] Fretwell, R., Kouvatsos, D.: LRD and SRD traffic: review of results and open issues for the batch renewal process. Performance Evaluation 48, 267–284 (2002)
[26] Jagerman, D.L., Melamed, B.: The transition and autocorrelation structure of tes processes part i: General Theory. Stochastic Models 8, 193–219 (1992)
[27] Erramilli, A., Singh, R.P., Pruth, P.: Chaotic maps as models of packet traffic. In: Labetoulle, J., Roberts, J.W. (eds.) The Fundamental Role of Teletraffic in the Evolution of telecommunications Networks, Proc. ITC-14, Antibes, Juan-les-Pins, France, June 1994, pp. 329–338. Eleseveir, Amstrerdam (1994)
[28] Maulik, K., Resnick, S.: Small and Large Time Scale Analysis of a Network Traffic Model. Queueing Systems 43, 221–250 (2003)
[29] Robert, S., Le Boudec, J.-Y.: New models for pseudo self-similar traffic. Performance Evaluation 30(1), 57–68 (1997)
[30] Courtios, P.J.: Decomposability, Queue and Computer Applications. Academic Press, New York (1977)
[31] Simon, H., Ando, A.: Aggregation of variables in dynamic systems. Econometrica 29, 111–139 (1961)
[32] Paxson, V.: Fast, Approximate Synthesis of Fractional Gaussian Noise for Generating Self-Similar Network Traffic. Computer Communications Review 27, 226–244 (1997)
[33] King, P.J.B., Larijani, H.A.: Queueing Consequences of Self-similar Traffic. In: Proc. Fourteenth UK Computer and Telecommunications Performance Engineering Workshop, Edinburgh, pp. 182–186 (1998)

[34] Larijani, H.A., King, P.J.B.: Effect of Packet Size on the Response Time in an Ethernet LAN with Self-Similar Traffic. In: Proc. IFIP ATM & IP 2000, Ilkley,West Yorkshire, U.K (2000)
[35] Larijani, H.A., King, P.J.B.: Response Time in an Ethernet LAN with Self-Similar Traffic. In: Proc. PGNET 2000, Liverpool (2000)
[36] Christensen, K.J., Molle, M.L., Yeger, B.: The design of a station-centric network model for evaluating changes to the IEEE 802.3 Ethernet standard. Simulation 72(1), 33–47 (1999)
[37] Neuts, M.F.: Matrix Geometric solutions in Stochastic Models: An Algorithmic Approach. John Hopkins University Press, Baltimore (1981)