Long Range Dependent Traffic Models

Eeva Nyberg
eeva.nyberg@hut.fi

25.10.99

Abstract

The report reviews papers on long-rang dependent (LRD) traffic models. Though many traffic measurements have shown the presence of LRD in packet networks, questions on the need of LRD traffic models are still posed. Thus, the issues are considered in this work. The report shows, on the other hand, that LRD may be an artifact of non-stationarity, or it may be ignored in the presence of strong short-range correlations. However, traffic measurement studies also show the need for LRD traffic models in order not to give overly optimistic performance predictions that result in underestimated network resource allocations. The papers discussed in this work have simultaneously investigated the effect of LRD on queuing performance and modeled packet traffic that exhibits moderate to strong LRD.

Contents

1 Introduction

Measurements on packet networks have shown that aggregate packet streams are statistically self-similar in nature. Long-range dependency (LRD) is one of many mathematical properties of self-similar traffic. Consider a process $Y = (Y_t : t = 0, 1, ...)$ where all Y_t have a common mean and common finite variance and the process has a covariance function independent of t, namely, $\gamma_k = \mathrm{Cov}(Y_t, Y_{t+k})$. Such a process is said to be covariance (or second-order or wide-sense) stationary. The auto-correlation function for this process is $\rho_k = \frac{\gamma_k}{\gamma_0}$. If we define the aggregate state $S_m = \sum_{k=1}^{m} Y_k$, it has variance of the form

$$\mathrm{Var}(S_m) = \sigma^2 \Big[m + 2 \sum_{k=1}^{m} (m-k)\rho_k \Big].$$

Let $v_m = \mathrm{Var}(S_m/m)$. Then if the process satisfies the following conditions:

1. $\sum_{k=1}^{\infty} \rho_k < \infty$
2. For large m, $v_m \approx \sigma^2/m$
3. For large m, $\mathrm{Var}(S_m) \approx \sigma^2[m^2 + o(m)]$

it is called short-range dependent. Examples of these processes include Markov chains and auto-regressive moving average (ARMA) processes of finite order.

On the other hand if it has the following properties:

1. $\sum_{k=1}^{\infty} \rho_k = \infty$
2. $mv_m \to \infty$ as $m \to \infty$
3. For large m, $\mathrm{Var}(S_m) \approx \sigma^2 m^{2-\alpha}$, $0 < \alpha < 1$

it is said to exhibit long-range dependence. Defining $H = 1 - \alpha/2$, where H is called the Hurst parameter, we obtain a measure of the degree of LRD; for short-range dependence, $H = 1/2$ and for long-range dependence, $1/2 < H < 1$.

Investigating the conditions listed, the distinction between the two processes can be made based on the sum of the auto-correlations or on the basis of the asymptotic rate of growth of the variance of S_m.

After shortly introducing the tests used to identify LRD in traffic measurements and the assumptions behind the tests in section 2, different traffic models used to characterize LRD-traffic will be discussed in section 3. In section 4 the impacts of LRD on traffic engineering, especially from the point of view of variable bit rate-video (VBR) and to some extent from the point of view of packet data in general, will be discussed. The report is concluded by section 5.

2 LRD vs. Non-Stationarity

Before introducing LRD traffic models, the assumptions behind some graphical and statistical methods for testing long-range dependence in a time series should be considered. The tests used include the variance-time analysis, the R/S analysis, the Index of Dispersion for Counts and the Periodogram-based analysis. Some of these tests were summarized in the report by Esa Hyytiä and they are also explained in [2], [6], and [9].

In [9] section 2.2, the stationarity assumptions behind these tests are discussed. It is pointed out that even when the tests give a Hurst parameter $H > 0.5$, which, if the tests are valid, should indicate long-range dependency, it is not necessarily clear that the traffic exhibits LRD, or that the Hurst parameter given is correct. The traffic can also be non-stationary. Recall that, see the introductory section, LRD was defined in the framework of stationarity. Therefore, if the stationarity assumption does not hold the Hurst parameters given by the test may be deceiving. Consequently, in the case of finite data sets the stationarity assumption must be tested.

In [9], the stationarity of MPEG3 video traffic streams was tested. The data points in the streams represent the total bit rate of one video frame. The test for stationarity, based on the Analysis of Variance (ANOVA), rejected the stationarity hypothesis. Therefore, it was concluded that the long-range dependence must be seen as an artifact of non-stationarity. The authors then proceed by introducing a new type of process, which is non-stationary, as a candidate for a traffic model, the shifting level process (SLP).

Beran et al. [2] however point out that LRD models seem to give conservative performance criteria and can be used to model highly non-stationary data when no parsimonious alternative models exist.

3 Source models

In [9] a review of LRD traffic models is given. The traffic models presented are Quasi-Markovian models, superposed ON/OFF models, and Fractional Brownian motion (FBM) models. In the earlier COST-report [8] models presented include rate processes with heavy-tailed period lengths, processes generated by deterministic chaotic maps, two component FBM-based traffic models, fractional ARIMA processes, and models based on Markov chains labeled pseudo self-similar models.

The source model for VBR broadcast video traffic is presented in section 3.1. Based on this model, the authors deduce in a later paper [6] that LRD is not a crucial property in determining the buffer behavior of VBR-video sources. The source model and the buffer model deserve therefore a closer look.

3.1 Models by Heyman and Lakshman

In [5] 11 long sequences of broadcast video traffic are analyzed and based on the results source models applicable for different sequences are given. They are however not able to conclude that a single model with physically meaningful parameters could be used for each sequence. The traffic measurements are done in such a way that the scene change frames can be identified. Distributions are found for the number of cells in scene change frames, the number of cells in frames immediately following scene change frames, the scene lengths, and the number of cells in frames within a scene.

From the 11 sequences measured it was deduced that the number of cells in scene change frames were uncorrelated and had different distributions (e.g. Weibull and Gamma). The model for the number of cells in the frame succeeding a scene change frame, Y_n, is given by linear regression,

$$Y_n = a + bX_n + \epsilon_n,$$

where X_n is the number of cells in the nth scene-change frame, a and b are the regression parameters, and $\epsilon_n \sim N(0,1)$. For most sequences, the scene lengths and the number of cells in frames within a scene followed Pareto, Weibull, or Gamma distributions.

Because the number of cells per frame in scene change frames was found to be uncorrelated between successive frames, the source model was only constructed for the number of cells per frame for frames within a scene.

For an example sequence, labeled film, the measurements indicated that cells per frame for frames within a scene had a Pareto-distribution,

$$\frac{\Gamma(\alpha+k)\lambda^\alpha x^{k-1}}{\Gamma(\alpha)\Gamma(k)(\lambda+k)^{k+\alpha}}.$$

Here $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt$.

The models used were the Markov chain and the discrete auto regressive (DAR) models.

Markov Chain Model

Let X_n be the number of cells in frame n and Y_n be the integer part of $X_n/50$. $(Y_n; n = 1, 2, ..., N)$ is modeled as a Markov chain with transition matrix $P = (p_{ij})$, estimated as

$$\hat{p}_{ij} = \frac{\text{number of transitions } i \text{ to } j}{\text{number of transitions out of } i}.$$

The Markov model is only used for sequences where the DAR model is not sufficiently accurate.

DAR

The transition matrix for the DAR(1) process is

$$P = \rho I + (1 + \rho)Q,$$

where ρ is the auto-correlation coefficient for intra-scene frames. Each row of Q consists of the Pareto (the distribution of cells per frame for frames within a scene is Pareto) probabilities $(f_0, f_1, ..., f_K, F_K^c)$, where $F_K^c = \sum_{k>K} f_k$ and K is the peak rate in cells per frame.

Simulation Results

The accuracy of the source models where tested by calculating the cell loss probabilities using simulations. The two sequences, film and isaura1, used were normal quality broadcast video, which had low mean rate and high peak-to-mean ratios. It was concluded that both models are accurate, but the DAR(1) model is preferred as it has only three parameters; mean, variance, and correlation. However for the other sequence the DAR(1) model was not accurate enough, giving overestimates for cell losses.

The paper concluded that a single model with a few meaningful physical parameters was not found for LRD traffic streams.

3.2 Models Given by Erramilli et al.

The authors performed queuing simulation experiments with traces of Ethernet LAN traffic and "shuffled" versions thereof. Different shuffled traces were used. The idea was to obtain traces with different correlation structures, but to preserve the arrival time distributions. In this way, the impact of LRD of the original traffic trace on queuing performance, dimensioning of buffers and determining usable capacity could be tested. If the impact of LRD were negligible as proposed by Heyman and Lakshman, then the shuffled trace preserving only the short-range dependency would give similar results to the original trace in the queuing simulations. This was not the case, and the authors therefore conclude that LRD has considerable impact on queuing performance and the use of models that incorporate LRD in a parsimonious matter (i.e. with as few parameters as possible) is justified.

In order to achieve parsimony in models the authors consider the features that contribute significantly to queuing performance. They consider conditions under which second order statistical descriptions of traffic processes are sufficient, discuss FBM, which parsimoniously captures LRD and demonstrate the effect of traffic shaping, e.g. limited buffer capacity, on LRD. These effects are discussed in section 4.2 in more detail. In this section, the FBM model is shortly introduced.

Fractional Brownian Motion

In standard Brownian motion models, the cumulative arrival process $A(t)$ is modeled by random fluctuations about a mean rate

$$A(t) = nt + \sqrt{an}Z(t), \tag{1}$$

where the process $Z(t)$ has independent Gaussian increments, and a is a peakedness term that describes the magnitude of fluctuations. In FBM the increments of $Z(t)$ are LRD and it has the following properties [8]:

1. $Z(t)$ has stationary increments,
2. $Z_0 = 0$ and $\mathrm{E}(Z_t) = 0$ for all t,
3. $\mathrm{E}(Z_t^2) = |t|^{2H}$ for all t,
4. Z_t has continuous sample paths
5. Z_t is Gaussian, i.e., all its finite-dimensional marginal distributions are Gaussian.

In the special case of $H = 1/2$, i.e. short-range dependency, Z_t is the standard Brownian motion.

The FBM model is valid, and complex short-range correlations can be ignored from a traffic engineering perspective, under the following conditions [3]:

1. time scales of interest in the queuing processes coincide with the scaling region
2. traffic is aggregated from a large number of independent users
3. effect of flow controls on any one user is negligible.

This implies that the long-range correlations can be parsimoniously modeled using the FBM model with three parameters; Hurst parameter H, magnitude parameter representing the strength of the fluctuations $\delta\lambda$, and the average rate $\bar{\lambda}$.

For an aggregation of n independent users, the traffic density in terms of time is

$$\lambda(t) = \sum_{i=1}^{n} \lambda_i(t),$$

where λ_i:s are uncorrelated for different users. The strength of the fluctuation is then the deviation from the mean of the traffic density

$$\delta\lambda = \lambda(t) - \bar{\lambda}.$$

For the FBM model used $\lambda_i(t) = 1$ giving $\lambda(t) = n$.

4 Queuing Performance

Most traffic models are used in order to get insight on queuing performance questions. If traffic exhibits LRD, many of the performance measures used change drastically. An example is the tail distribution of an infinite system, which changes from an exponential distribution to a Weibull distribution resulting in a heavier tail and therefore underestimation of loss probabilities.

4.1 Implications of LRD to VBR Traffic Engineering as Seen by Heyman and Lakshman

Heyman and Lakshman use the Markov chain model shown earlier to determine buffer occupancy, and conclude that LRD models are not needed when considering VBR traffic.

Buffer Model

The buffer model is a generic model. The buffer has capacity c and receives inputs at deterministic times. The time scale is chosen to match the application. If X_i is the number of arrivals at time T_i and d_i the number of items processed during $[T_i, T_i + 1]$ then the buffer content V_i at the end of the ith interval is given by

$$V_i = \min\{(V_{i-1} + X_i - d_i)\ +, c\}, i = 1, 2, \dots\ ,$$

where it is assumed that $V_0 = 0$.

Applied to video teleconferencing the time intervals are taken to have length equal to time between frames. The buffer model is solved for the infinite buffer case $c = \infty$. For the finite buffer case, the solution is only given for the first two intervals.

They conclude that the effect of LRD are significant only if LRD causes the busy periods to be long enough for the long lags to come into play. They argue further that since VBR services carrying video traffic will be delay sensitive (to avoid jitter) and sensitive to cell losses (to avoid picture degradation), the traffic intensity for these services will not be large. Therefore the busy periods for VBR traffic are short and the resetting effect should demonstrate itself strongly in practical operating regions. By the resetting effect the authors mean that only those sums (of $X_i s$) taken within a busy period affect the buffer size. An example of the resetting effect is demonstrated for the finite buffer case. The solution of V_i, for $i = 1, 2$ is

$$V_1 = \begin{cases} 0 & \text{if } Y_1 < 0 \\ Y_1 & \text{if } 0 \leq Y_1 \leq c \\ c & \text{if } Y_1 > c \end{cases}$$

$$V_2 = \begin{cases} 0 & \text{if } Y_2 < 0 \\ Y_2 & \text{if } 0 \leq Y_2 \leq c \\ c & \text{if } Y_2 > c \end{cases} \quad \text{, Case 1: } Y_1 < 0$$

$$V_2 = \begin{cases} 0 & \text{if } Y_2 < -Y_1 \\ Y_1 + Y_2 & \text{if } -Y_1 \leq Y_1 \leq c - Y_1 \\ c & \text{if } Y_2 > c - Y_1 \end{cases} \quad \text{, Case 2: } 0 \leq Y_1 \leq c$$

$$V_2 = \begin{cases} 0 & \text{if } Y_2 < c \\ c + Y_2 & \text{if } -c \leq Y_1 \leq 0 \\ c & \text{if } Y_2 > 0 \end{cases} \quad \text{, Case 3: } c < Y_1$$

Here $Y_i = X_i - d$. Case 1 demonstrates the resetting effect, as the buffer is empty at the start of the second interval. Case 3 demonstrates what the authors call the truncating effect of finite buffers, an enhancement of the resetting effect. The effect of Y_1 on V_2 is that $Y_1 > c$, but how large is irrelevant.

They also argue that as the processing rate d is increased the cell loss rate decreases and the busy periods are stochastically shorter. In such a case, the truncating effect of finite buffers gets stronger when the cell loss rate gets smaller. They therefore state that if experiments could be done with smaller cell-loss rates, the accuracy of Markov Chain models would be better than the accuracy achieved at the time of writing.

The simulation results show that the LRD is unimportant when there is strong short-range dependence and the Hurst parameter is not very large ($H < 0.7$) as was the case for the video conference data sets examined. The entertainment video sets, film and isaura1, had strong short-range dependencies, but Hurst parameters larger than 0.9. The truncating effect was strong enough so that the Markov models correctly estimated the cell loss rate. However, the mean buffer size was not estimated correctly. The authors however explain that the traffic intensity of the trace is above practical values, and when the buffer size is reduced to a practical value, the models estimate the mean buffer size accurately.

Note that Heyman and Lakshman were not able to model the sequences, film and isaura1, in a parsimonious matter with the DAR(1) model, contrary to the earlier video teleconference models given in [4].

4.2 Queuing Analysis with LRD Packet Traffic as Seen by Erramilli et al.

The FBM model given by Norros gives the asymptotic lower bound for the probability that the queue length v exceeds x [7],

$$P(V > x) \sim e^{-cx^{2-2H}}, \tag{2}$$

where

$$c = \frac{n^{2H-1}(\frac{1-\rho}{\rho})^{2H}}{2a} \cdot \left[\Big(\frac{1-H}{H}\Big)^{H} + \Big(\frac{H}{1-H}\Big)^{1-H}\right]^2.$$

Here ρ is the utilization, that is the ratio of $\bar{\lambda}$ to the processor speed.

The simulation results show that the Weibull-like asymptotic queue length curve given by the FBM model coincides with the original traffic trace driven simulations.

The need for LRD models can be further demonstrated by comparing equation 2 to the queue length probability for large x given by Markov models

$$P(V > x) \sim e^{-\eta x}. \tag{3}$$

Effective bandwidth-based admission-control algorithms using 3 will underestimate the target small cell blocking probabilities by several orders of magnitude when, in fact the sources exhibit even only moderate long-range dependence [2].

The claim made by Heyman and Lakshman in [6] on the buffering action of a queue is challenged by Erramilli et al. They consider the effect that traffic shaping, by buffering action or rate controlling, has on LRD. They conclude that at best controlling LRD by traffic shaping transfers the buffer requirements from within the network to the edge of the network, without an improvement on the performance. Additional issues from engineering implications of LRD are discussed in [3]. These include the effect of finite buffers, the physical basis for LRD, the statistical multiplexing gains, VBR video traffic, and estimating model parameters. Due to the claims made by Heyman and Lakshman, the finite buffer size and VBR video traffic deserve a closer look.

Finite Buffer Size

The functional form of the queue length distribution for the finite buffer systems remains Weibull. The difference is the well-known dominance of the infinite buffer queue length distribution over its finite buffer counterpart. Although a finite buffer will not show a sharp degradation in delay performance observed with infinite buffer systems, the packet loss increases in the case of the finite buffer. Many data applications are tolerant to delays, but highly sensitive to losses. The use of the infinite buffer regime is therefore of practical interest and the LRD has to be therefore modeled. Note that in [6] the Markov chain models were successful only when the traffic intensities were reduced. Though they argued that these levels were those found in practice, Erramilli et al. argue that it may not be economical to keep packet losses acceptable by operating the network at low utilizations so that the correlation never impact the performance of the queuing system.

VBR Video Traffic

Erramilli et al. agree that VBR traffic that has strong short-term correlations in the data make it only "asymptotically" self-similar. In contrast, data traffic, they argue, can be modeled by exactly self-similar processes. A parsimonious model that models both the short-range and long-range correlation structures would be ideal. On suggestion given in [8] is to add a Brownian motion component W_t to the FBM traffic model in equation 1,

$$A_t = nt + \sqrt{b}W_t + \sqrt{a}Z_t. \tag{4}$$

4.3 Superposition of MMPPs

In addition to the models discussed so far, Andersen and Nielsen [1] have taken the Markovian approach, in a more recent paper, for modeling LRD packet traffic. The models consist of superpositions of two state Markov Modulated Poisson Processes (MMPPs). They show that a superposition of four MMPPs suffices to model second-order self-similar behavior. They apply these models to counting and interarrival time processes. However, the drawback of this approach is that the models do not appear sufficient to for accurate prediction of queuing behavior.

5 Conclusions

The papers reviewed in this report have tried simultaneously to investigate the effect of LRD on queuing performance and to model packet traffic that exhibits moderate to strong LRD.

Heyman and Lakshman concentrated on VBR traffic that exhibited moderate long-range dependencies. They were able to construct parsimonious models for traffic with strong short-range correlations. Thy were also able to conclude that LRD is not a crucial property on queuing performance issues when dealing with VBR traffic with moderate LRD and strong short-range dependency.

Erramilli et al. considered data traffic in general and showed using shuffled traces that LRD is a traffic characteristic that cannot be ignored in traffic engineering. They however point out that FBM models, which only model LRD, may not be satisfactory for modeling packet traffic that exhibits strong short-range correlation. It therefore seems that the use of models created by Heyman and Lakshman are useful in this context.

Another proposed model is that given in equation 4, where both the long and short-range correlations are taken into account in the traffic model. However, the question of parsimony cannot be neglected, and the use of either a short-range or a long-range model may be appropriate when only one of them dominates strongly.

References

[1] A.T. Andersen and B.F: Nielsen, A Markovian approach for modeling packet traffic with long-range dependence, IEEE JSAC 16, 5,719-732, June 1998

[2] J. Beran, R. Sherman, M.S. Taqqu, and W. Willinger, Long-range dependence in variable-bit-rate video traffic, IEEE TCOM, 43, 2/3/4,1566-1579, February/March/April 1995

[3] A. Erramilli, O. Narayan, and W. Willinger, Experimental queuing analysis with long-range dependent packet traffic, IEEE/ACM TNET 4,2, 209-223, April 1996

[4] D.P. Heyman, A-Tabatabai, and T.V. Lakshman, Statistical analysis and simulation study of video teleconference traffic in ATM networks, IEEE Trans. Circuits. Syst. Video. Technol., 49-59, March 1992

[5] D.P. Heyman and T.V. Lakshman, Source models for VBR broadcast-video traffic, IEEE/ACM TNET 4, 1, 40-48, February 1996

[6] D.P. Heyman and T.V. Lakshman, What are the implications of long-range dependence for VBR-video traffic engineering?, IEEE/ACMTNET 4, 3, 301-317, June 1996

[7] I. Norros, A storage model with self-similar input, Queuing Systems, Vol. 16, 387-396, 19994

[8] J. Roberts, U. Mocci, and J. Virtamo (Eds.), Broadband Network Teletraffic, COST242 Final Report, Springer, Berlin, Sect. 13.4. (Long range dependent traffic models), 1996

[9] S. Molnar and I. Marica (Eds.), Source characterization in broadband networks, Proc. COST257 Mid-term Seminar, Vilamoura, Portugal, Section 2 (Fractal traffic characterization), January 1999