Queueing analysis with traffic models based on deterministic dynamical systems

Venkat Anantharam
EECS Department
University of California
Berkeley, CA 94720
ananth@vyasa.eecs.berkeley.edu

Abstract

Measurements of communication traffic from a wide variety of sources have suggested that correlations persist in such sources over fairly long time scales. This has called into question the use of traditional traffic models for performance analysis, and sparked an interest in studying novel kinds of traffic models. In particular, a suggestion has been made to build models based on expanding deterministic dynamical systems. Deterministic dynamical systems can exhibit chaotic behavior, which has many of the features of statistical behavior, and can indeed be studied rigorously using probabilistic techniques. We make some remarks regarding the analysis of queues driven by such traffic models.

1 Introduction

Statistical analyses of measurements of communication traffic from a wide variety of sources have suggested that correlations persist in such sources over fairly long time scales; for examples of such measurements, see [3], [16], and [21]. This has called into question the use of traditional traffic models for the performance analysis of communication networks, and has sparked considerable research on the use of "long-range dependent" traffic models for performance analysis; for examples see [1], [2], [14], [17], [19], and [20]. A feature of most of these works is the observation that the tail behavior of queues with long-range dependent inputs decays much slower than exponentially, either as a Weibull or according to a power law, depending on the model. The presence of a qualitative difference in queueing behavior with the traffic measured in [16] as compared to that predicted by conventional models is also supported by the experimental analysis of [11]. The persistence of long-range dependence in packet traffic even after it is regulated by simple flow control schemes

*Research supported by NSF NCR 9422513.

is demonstrated in [24]. A bibliography of work in this area as of the middle of 1996 is available in [26].

It should be pointed out that this area is quite active, and the jury is still out on the interpretation and significance of these results. There are at least two general grounds for caution. The first has to do with the very existence of "long-range dependence" (a nice discussion of a somewhat analogous earlier controversy in the field of geology is in [13]). For instance, the analysis of the Ethernet measurements in [16] is primarily done using the R/S statistic, variance-time plots, and the periodogram. The R/S statistic can be highly sensitive to trends in the observations; for instance, a special case of the main result of [4] is that the sum of an independent and identically distributed (i.i.d.) sequence of random variables with finite mean and variance and a deterministic trend that is vanishingly small and goes to zero asymptotically in time according to a power law, can result in the appearance of a nontrivial (not equal to 1/2) Hurst exponent when analyzed using the R/S statistic (for a precise formal theorem, see [4]). The behavior of variance-time plots for the Bellcore measurements of [16] under different kinds of nonstationarity is discussed in [8], and also in [23] (in the latter it is concluded that the measurements support the the hypothesis of long-range dependence fairly well). Certainly, the variance-time plots for the AUG89.MB measurements of [16] (plotted in Fig. 5 of [16]) would seem to support the existence of correlation up to a scale of about 100 seconds, under the hypothesis of stationarity, as a crude back of the envelope calculation of empirical variances for a stationary process model will show (see the appendix). This brings us to the second ground for caution, which is whether traditional traffic models with sufficiently large depth of correlation might already serve adequately for buffer dimensioning and network design purposes. Recent works that sound this note of caution include [12] and [25].

Whatever side the coin may fall on in this controversy, the measurements and statistical analyses of works such as [3], [16], and [21] are fascinating, and beg for some kind of physical explanation, whether it be due to trends, the nature of the underlying protocols, or the existence of "true" "long-range dependence". At least, such a physical understanding would aid in the construction of parsimonious and realistic traffic models, be they "traditional" or "long-range-dependent". One attempt at such a physical explanation is in [27], where it is argued via statistical analysis of the measurements in [16] that they are consistent with infinite variance of the load at the level of individual sources. It is proved in [22] that the superposition of ON/OFF sources whose ON-periods or OFF-periods have infinite variance results in an aggregate network traffic that is long-range dependent.

In this note our purpose is to make a couple of remarks in connection with another intriguing idea for arriving at physical descriptions of network communication traffic, suggested in [10].

2 Deterministic dynamical systems

The proposal in [10] is to model a communication traffic source (such as an Ethernet LAN, or a VBR video source) by means of a deterministic nonlinear transformation

$x_{n+1} = S(x_n)$, taking values in some state space X. The traffic source is modeled as having an activity level that depends on the current state.

Let us first remark that if X is allowed to be sufficient general, the restriction to deterministic S is not a big one. Indeed, the theory of deterministic chaos, see for instance [5] or [15], tells us that quite complicated statistical behavior can be expressed by such deterministic maps. For instance, consider the transformation on the unit interval $f : [0,1] \mapsto [0,1]$ given by

$$S(x) = \begin{cases} \frac{x}{p} & \text{if } 0 \leq x \leq p \\ 1 - \frac{x-p}{1-p} & \text{if } p \leq x \leq 1 \ . \end{cases}$$

Then, starting from Lebesgue almost any initial condition x_0, the empirical distribution of the sequence of iterates $\{S^n(x_0), n \geq 0\}$ will converge to Lebesgue measure on $[0,1]$ (in the weak topology of convergence of measures). Thus, there appears to be a stationary situation in which the state is distributed according to Lebesgue measure. Suppose the activity level of the source is

$$a(x) = \begin{cases} a & \text{if } 0 \leq x \leq p \\ 0 & \text{if } p \leq x \leq 1 \ . \end{cases}$$

If one were to start with the initial state distributed according to Lebesgue measure, the sequence of iterates $\{a(x_n), n \geq 0\}$ of the source values will be a sequence of i.i.d. $\{0, a\}$ valued Bernoulli random variables with probability of being a equal to p.

The point just made is that deterministic maps can be used to model stochastic processes also. Thus, by choosing X and the transformation S appropriately it should be possible to directly model the traffic generated by fairly complicated protocols and systems, in a directly tangible and physically meaningful way, while at the same time retaining the flexibility to incorporate traditional stochastic processes into the models. Another argument for the potential interest in such deterministic models is that there appears to be a significant amount of determinism in the structure of communication traffic - in Fig 1. of [10] the successive interarrival times from the Ethernet measurements of [16] is plotted, and visual inspection shows what appears to be considerable deterministic structure in this plot.

We now proceed to make a couple of remarks about the queueing behavior of queues driven by such traffic models.

3 Large deviations

We first record a special case of a result of [6], see Thm 3.9 (ii) of that paper. Let $\{a_n, n \geq 0\}$ be a stationary and ergodic sequence of nonnegative real valued random variables, interpreted as the sequence of arrivals into a single server queue, which can serve an amount c per unit time. Thus, the queue size evolves according to the equation

$$q_{n+1} = (q_n + a_n - c)^+ \ .$$

We assume that the stability condition $E[a_n] < c$ holds. From the result of [18], we know that the queue has a unique stationary distribution. Let q_∞ be a random variable with this distribution. We assume that for any θ, $0 < \theta < \infty$, the limit

$$a^*(\theta) \triangleq \frac{1}{n} \log E[e^{\theta \sum_{t=0}^{n-1} a_t}] \tag{1}$$

exists. Let us define

$$\theta^* \triangleq \sup\{\theta \ : \ a^*(\theta) < c\} \ .$$

Then the claim of [6] is that we have

$$\lim_{u\to\infty} \frac{-\log P(q_\infty \geq u)}{u} = \theta^* \ .$$

In this result, note that, since $\theta^* > 0$, the tail probability of the stationary queue size decreases exponentially.

We next note a special case of a result recorded in [7], see Theorem 6.4.4. and pg. 261 of that book. Given an $\mathbf{R}^K$- valued ψ-mixing sequence of random variables $\{Y_n, n \geq 0\}$, the sequence of empirical averages $\{S_n, n \geq 0\}$, where

$$S_0 \triangleq 0, \quad S_n \triangleq \frac{1}{n} \sum_{t=0}^{n-1} Y_t \ ,$$

obeys a large deviations principle with good convex rate function $\Lambda^*(\cdot)$, which is the Legendre transform of the function

$$\Lambda(\lambda) = \lim_{n\to\infty} \frac{1}{n} \log E[e^{n<\lambda, S_n>}] \ , \tag{2}$$

defined for $\lambda \in \mathbf{R}^K$, $\lambda = (\lambda_1, \ldots, \lambda_K)^T$. Here $< \lambda, S_n >$ denotes the inner product of vectors in $\mathbf{R}^K$. In particular, this result says that the limit in equation (2) exists.

4 Main remarks

Consider a deterministic dynamical system, given by a map $S : X \mapsto X$, where $(X, \mathcal{F}, \mu)$ is a measure space. A priori, the measure μ has nothing to do with the dynamics of S; it serves as a reference measure to permit talking about densities, and, as is customary, we will assume that S is measurable, and *nonsingular* with respect to μ, i.e. the induced measure $S_*(\mu)$ (which is also often written $\mu(S^{-1})$) is absolutely continuous with respect to μ.

We also assume given a partition of X into K measurable subsets $I_1, \ldots, I_K$, real numbers $a_1, \ldots, a_K$ (which need not be distinct), and define the function $a : X \mapsto \{a_1, \ldots, a_K\}$ by

$$a(x) = \sum_{k=1}^{K} a_k 1_{I_k}(x) \ ,$$

where $1_{I_k}(x)$ is the indicator function of I_k.

Suppose we start the dynamical system with a distribution that is absolutely continuous with respect to μ and can therefore be written as $f\mu$ for some density $f \in L^1(\mu)$ (i.e. $f \geq 0$ and $\int_X f(x)\mu(dx) = 1$). Then the next state will also have a distribution that is absolutely continuous with respect to μ (because S is nonsingular with respect to μ), so there is an operator P on $L^1(\mu)$ such that $P(f)$ is the density of this next state with respect to μ. P is called the *Frobenius - Perron operator* of the transformation.

Under a wide range of conditions it is known that, for arbitrary initial density f, the sequence of iterates $\{P^n(f), n \geq 0\}$ will converge to a limit f^* such that $P(f^*) = f^*$. For details regarding the sense in which convergence takes place, and situations in which it is know to take place, see [5] and [15]. There is a vast literature this topice, with conditions known that cover several nonlinear maps with quite nontrivial behavior. For maps of the unit interval $[0, 1]$, for instance, a commonly cited example of such a result is recorded in Theorem 6.2.1 of [15]. Thus, the state of the dynamical system appears to converge, in some appropriate sense, to a random state with distribution $f^*\mu$. If one starts the dynamical system with this distribution, then the sequence of arrivals $\{a_n, n \geq 0\}$ will be a stationary ergodic process. It is certainly the case that under a wide range of conditions, this process will be ψ-mixing; indeed, the underlying dynamical system itself can be shown to be strong mixing in many cases; for instance, *exactness* of the transformation, which is often quite easy to see, implies mixing.

Combining these observations with the results stated in the preceding section leads to our first remark : If one constructs such an expanding discrete dynamical system model, and an arrival process derived from such a model, then, in the (non-pathological) situation where a stationary distribution exists for the dynamical system, one should typically expect exponential decay of the tails of the stationary queue size in queues driven by the arrivals. It suffices to apply the second result of section 3 to the sequence of $\mathbf{R}^K$ valued random variables

$$\{(1_{I_1}(x_n), \ldots, 1_{I_K}(x_n)),\ n \geq 0\}\ ,$$

to conclude the existence of the limit in equation (1).

We next observe that positive recurrent discrete time finite state Markov chain arrivals are quite easily constructed along the lines of the general scheme we are discussing. To get a stationary sequence $\{a_n, n \geq 0\}$ taking values in $\{a_1, \ldots, a_K\}$ with stationary distribution $(\pi_k, 1 \leq k \leq K)$ and transition probability matrix $(p_{ij}, 1 \leq i, j \leq K)$, take X to be the unit interval $[0, 1]$, $\mathcal{F}$ to be the Borel σ-field, and μ to be Lebesgue measure. Take $I_1, \ldots, I_K$ to be intervals of lengths $\pi_1, \ldots, \pi_K$ respectively, and further partition each I_k into intervals I_{kl}, $1 \leq l \leq K$, of lengths $\pi_k p_{kl}$ respectively. Consider the deterministic transformation

$$S : [0, 1] \mapsto [0, 1]$$

which is *piecewise linear* and maps the interval I_{kl} onto the interval I_l (so that it has slope $\frac{\pi_l}{\pi_k p_{kl}}$ on this interval. Except for the quibble that a state of the Markov chain may lead to another state of the chain with probability 1, which situation can also be easily handled, this is an example of a piecewise linear (expanding) Markov transformation, in the sense of Chapter 9 of [5], because the slopes are all strictly bigger than 1; it is easily seen to leave Lebesgue measure invariant (so the invariant density is 1). The sequence of arrivals $\{a_n, n \geq 0\}$, when we start the dynamical system with Lebesgue measure, will be a Markov chain with initial distribution $(\pi_k, 1 \leq k \leq K)$ and transition probability matrix $(p_{ij}, 1 \leq i, j \leq K)$.

Our second remark is the following : Suppose we are given any piecewise expanding transformation of the interval (for the definition, see page 85 of [5]). These are among the most studied transformations in the literature, and for maps of the interval an expanding condition is required for most results regarding the existence of a stationary distribution. [1] Then one can approximate the transformation by a piecewise linear (expanding) Markov transformation. The arrival process corresponding to this approximation, in stationarity, is a function of a stationary finite state Markov chain. The limit of equation (1) can now be *explicitly* written down in terms of the slopes of the approximation to the transformation. This is a basic result that goes back to the seminal works of Donsker and Varadhan for the large deviations of the empirical distribution of Markov processes; for a convenient reference for such results, see [9]. Thus, according to the result of [6], the tail probabilities of a single server queue driven by such an arrival process will decay exponentially (assuming stability), and the rate of decay of the tail can be explicitly computed in terms of the parameters of the approximating Markov transformation.

Note that our second remark can be generalized to transformations of spaces other than the interval - all one needs is that it should be possible to construct a finite state symbolic dynamical scheme that approximates the given transformation.

5 Conclusion

Deterministic dynamical systems offer a promising (or at least suggestive !) modeling paradigm for the communication traffic processes generated by systems executing complex protocols, which might also be composed of several interacting components. It is important to realize that one has *not* lost modeling flexibility by restricting to deterministic transformations, because stochastic processes can also be modeled by this paradigm. What one appears to gain is a direct physical approach to building models based on an explicit description of the underlying system dynamics. We have pointed out that when we consider a queue fed by arrivals derived from such a system, we should typically expect to see exponentially decaying tails for the stationary queue size. Further, by chosing suitable approximations for the transformation describing the system, we can write explicit formulas for the rate of decay of the tail probabilities of the stationary queue size.

[1] The models proposed in [10] are also of this type, although the focus there is on conjectures regarding the situation corresponding to "intermittency" in physics, where the expanding condition fails to hold at one point.

6 Appendix

Let $X_1, \ldots, X_N$ be a sequence of conditionally independent Poisson random samples, conditioned on means $\mu_1, \ldots, \mu_N$ respectively. Assume $\mu_n = \lambda + \gamma_n$, $1 \leq n \leq N$. This is a simple model for a sequence of packet or byte counts with fluctuating mean. The samples are aggregated at level m to get

$$X_j^{(m)} = X_{(j-1)m+1} + \ldots + X_{jm} \ , \quad 1 \leq j \leq M \ ,$$

where $1 \leq m \leq N$, m divides N, and M denotes N/m. The variance time plot is the plot of the logarithm of the empirical variance of the m-aggregated sequence $X_1^{(m)}, \ldots, X_M^{(m)}$ against $\log m$.

Write Γ_n for $\sum_{t=1}^{n} \gamma_t$. The conditional expectation of the empirical variance of the m-aggregated sequence is seen to be

$$\frac{\lambda + \Gamma_N/N}{m} - \frac{\lambda + \Gamma_N/N}{N} + \frac{1}{Nm} \sum_{j=1}^{M} (\Gamma_{jm} - \Gamma_{(j-1)m})^2 - \frac{{\Gamma_N}^2}{N^2} \ .$$

Thus, loosely speaking, if $\Gamma_N = o(N)$, and, for most j, $(\Gamma_{jm} - \Gamma_{(j-1)m})^2 \sim m^{2H}$ over the range $m_0 \leq m \leq \frac{N}{m_0}$, then the variance time plot will exhibit a slope of $1 - 2H$ over this range of m. On the basis of this crude calculation the variance time plot of Fig. 5(b) of [16] could be taken to indicate correlations on the scale of about 100 seconds for the AUG89.MB measurements described there, but do not necessarily indicate correlations on longer time scales than this.

References

[1] Anantharam, V. (1995). On the Sojourn Time of Sessions at an ATM Buffer with Long-Range Dependent Input Traffic. *Proceedings of the 34rd IEEE Conference on Decision and Control*, New Orleans, Louisiana, December 13 -15, 1995, Volume 1, pp. 859 -864.

[2] Anantharam, V. (1996). "Networks of Queues with Long-Range Dependent Traffic Streams", *Stochastic Networks: Stability and Rare Events*, edited by Paul Glasserman, Karl Sigman, and David Yao, Lecture Notes in Statistics, Vol. 117, Springer Verlag, pp. 237 - 256.

[3] Beran, J., Sherman, R., Taqqu, M.S., and Willinger, W. (1995). Long-range dependence in variable-bit-rate video traffic. *IEEE Transactions on Communications*, Feb.-April 1995, vol.43, Nos. 2-4, pt.3, pp. 1566 -1579.

[4] Bhattacharya, R. N., Gupta, V. K., and Waymire, E. (1983). The Hurst Effect under Trends. *Journal of Applied Probability*, Vol. 20, pp. 649 -662.

[5] Boyarsky, A. and Góra, P. (1997). *Laws of Chaos : Invariant Measures and Dynamical Systems in One Dimension.* Birkhäuser, Boston.

[6] Chang, C. S. (1994). Stability, Queue Length, and Delay of Deterministic and Stochastic Networks. *IEEE Transactions on Automatic Control*, Vol. 39, No. 5, pp. 913 -931.

[7] Dembo, A. and Zeitouni, O. (1993). *Large Deviation Techniques and Applications*. Jones and Bartlett Publishers, Boston.

[8] Duffield, N.G., Lewis, J.T., O'Connell, N., Russell, R., and Toomey, F. (1994). Statistical issues raised by the Bellcore data. *IEE Eleventh UK Teletraffic Symposium. Performance Engineering in Telecommunications Networks*, Cambridge, UK, Vol 1 pp. 1-8.

[9] Ellis, R. S. (1988). Large Deviation for the Empirical Measure of a Markov Chain with an Application to the Multivariate Empirical Measure. *Annals of Probability*, Vol. 16, pp. 1496 -1508.

[10] Erramilli, A. and Singh, R. P. (1995). An application of deterministic chaotic maps to model packet traffic. *Queueing Systems : Theory and Applications*, Vol 20, pp. 171 -206.

[11] Erramilli, A., Narayan, O., and Willinger, W. (1996). Experimental queueing analysis with long-range dependent packet traffic. *IEEE/ACM Transactions on Networking*, April 1996, vol.4, no.2, pp. 209-23.

[12] Heyman, D.P. and Lakshman, T.V. (1996). What are the implications of long-range dependence for VBR-video traffic engineering?. *IEEE/ACM Transactions on Networking*, June 1996, vol.4, no.3, pp. 301 -317.

[13] Klemes, V. (1974). The Hurst phenomenon : a puzzle ? *Water Resources Research*, Vol. 10, pp. 675 -688.

[14] Konstantopoulos, T. and Lin, S-J. (1996). Fractional Brownian Motion and Lévy Motions as Limits of Stochastic Traffic Models. *Proceedings of the 34th Annual Allerton Conference on Communications, Control, and Computing*, Urbana, Illinois, pp. 913 -922.

[15] Lasota, A. and Mackey, M. C. (1994). *Chaos, Fractals, and Noise : Stochastic Aspects of Dynamics*. Springer Verlag, New York.

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

[17] Likhanov, N., Tsybakov, B. and Georganas, N. D. (1995). Analysis of an ATM Buffer with Self-Similar ("Fractal") Input Traffic. *Proceedings of the 14th Annual IEEE Infocom*, pp. 985-992.

[18] Loynes, R. M. (1962). The Stability of Queues with Non-Independent Interarrival and Service Times. *Proceedings of the Cambridge Philosophical Society*, Vol. 58, pp. 497-520.

[19] Norros, I. (1994). A storage model with self-similar input. *Queueing systems : Theory and applications*, Vol. 16, pp. 387 -396.

[20] Parulekar, M. and Makowski, A. M. (1996). Tail Probabilities for a Multiplexer with Self-similar Traffic. *Proceedings of the IEEE INFOCOM 1996*, pp. 1452 -1459.

[21] Paxson, V. and Floyd, S. (1995). Wide Area Traffic: The Failure of Poisson Modeling. *IEEE/ACM Transactions on Networking*, Vol. 3, No. 3, pp. 226-244.

[22] Taqqu, M.S., Willinger, W., and Sherman, R. (1997). Proof of a fundamental result in self-similar traffic modeling. *Computer Communication Review*, April 1997, vol.27, no.2, pp. 5-23.

[23] Teverovsky, V. and Taqqu, M. (1995). Testing for Long-Range-Dependence in the Presence of Shifting Means or a Slowly Declining Trend, using a Variance-type Estimator. *Journal of Time Series Analysis*, Vol. 18, No. 3, pp. 270 -304.

[24] Vamvakos, S. and Anantharam, V. (1998). On the Departure Process of a Leaky Bucket System with Long-Range Dependent Input Traffic. *Queueing Systems : Theory and Applications*, 1998. to appear. Available as *UCB/ERL Memorandum*, No. M96/51 and *UCB/ERL Memorandum*, No. M96/70, Electronics Research Laboratory, University of California, Berkeley.

[25] Venkateshwara Rao, B., Krishnan, K.R., and Heyman, D.P. (1996). Performance of finite-buffer queues under traffic with long-range dependence. *Proceeding of the IEEE GLOBECOM 1996*, London, UK, vol.1., pp. 607 -611.

[26] Willinger, W., Taqqu, M.S., and Erramilli, A. (1996). A Bibliographical Guide to Self-similar Traffic and Performance Modeling for Modern High-speed Networks. In *Stochastic Networks : Theory and Applications*, edited by F. Kelly, S. Zachary, and I. Ziedins, Oxford University Press, Oxford, pp. 339 -366.

[27] Willinger, W., Taqqu, M.S., Sherman, R., and Wilson, D.V. (1997). Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level. *IEEE/ACM Transactions on Networking,* Feb. 1997, vol.5, no.1, pp. 71 -86.