Proceedings of the 34th
Conference on Decision & Control
New Orleans, LA - December 1995

On the Sojourn Time of Sessions at an ATM Buffer with Long-Range Dependent Input Traffic

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

Abstract [1]

Several recent empirical studies of traffic to be carried by broadband networks have established the importance of studying queueing behaviour with traffic models having long-range dependence. We consider discrete time models for an ATM buffer fed by a class of long-range dependent arrival processes that includes the asymptotically self-similar arrival processes of the kind recently proposed by Likhanov et al. [8]. An arrival model in this class is parametrized by a session generation rate $\lambda > 0$, a cell generation rate per session $R > 0$, and a session duration random variable τ which is positive integer valued with finite mean and infinite variance and has regularly varying tail, i. e. $P(\tau > t) \sim t^{-\alpha} L(t)$ where $1 < \alpha < 2$ and $L(\cdot)$ is a slowly varying function. The service is modelled by a single parameter, the service rate $s > 0$. The stability condition $s > R\lambda E\tau$ is assumed to hold. We study the asymptotics of the tail of the stationary sojourn time distribution of a typical session.

1 Introduction

Several recent empirical studies of traffic to be carried by broadband networks have established the importance of studying queueing behaviour with traffic models having long-range dependence. For instance, Leland et al. [7] have demonstrated the self-similar nature of Ethernet traffic by a statistical analysis of Ethernet traffic measurements at Bell-Core; Beran et al. [2] have demonstrated long-range dependence in samples of variable bit rate video traffic generated by a number of different codecs; and Paxson and Floyd [11] have concluded the presence of long-range dependence in TELNET and other wide area network traffic.

Queueing models are traditionally used to analyze the behaviour of buffered systems carrying traffic, such as communication networks. In order to appropriately design these systems the distribution of the traffic backlog in the buffers and of the delay incurred by traffic are of interest. Motivated by the empirical evidence for long-range dependence in network traffic, a number of studies have recently appeared that investigate single-server queues with long-range dependent arrival models. These include the works of Norros [10] (discussed further by Duffield and O'Connell [5]) who uses a fractional Brownian arrival model; Resnick and Samorodnitsky [12] who use an arrival model derived from a stable integral; and Likhanov et al. [8] who use an asymptotically self-similar discrete time arrival model (which also appears in the survey of Cox [4, page 68]).

In this note we consider a single-server queue driven by a class of discrete time long-range dependent arrival models that includes, as a special case, the asymptotically self-similar arrival processes of the kind proposed in [8]. An arrival model in this class is parametrized by a session generation rate $\lambda > 0$, a cell gen-

[1] Research supported in part by NSF grant NCR 94-22513 and by the AT & T Foundation

eration rate per session $R > 0$, and a session duration random variable τ which is positive integer valued with finite mean and infinite variance and has regularly varying tail, i. e. $P(\tau > t) \sim t^{-\alpha}L(t)$ where $1 < \alpha < 2$ and $L(\cdot)$ is a slowly varying function. The service is modelled by a single parameter, the service rate $s > 0$. The stability condition $s > R\lambda E\tau$ is assumed to hold.

Suppose the sessions are served in FIFO order of arrival. When the additional condition $s \leq R$ holds no session will enter service prior to the completion of all work brought in by preceding sessions. We first observe that in this regime classical results in queueing theory allow one to explicitly determine the asymptotics of the tail of the stationary sojourn time distribution of a typical session. We next investigate the case $s > R$. In this situation we give an asymptotic upper bound for the tail of the stationary sojourn time distribution of a typical session.

Without loss of generality, we will set $R = 1$ throughout the subsequent discussion. As usual $\mathbf{N}$ and $\mathbf{Z}$ denote the set of natural numbers and the set of integers respectively. Σ^* will denote a random variable with the distribution of the stationary sojourn time of a typical session. Thus the asymptotic behaviour of $P(\Sigma^* > x)$ in x is the topic of interest in this note.

2 Preliminaries

We consider a class of discrete time models for arrivals into an ATM buffer. The duration between time l and $l+1$ is called *slot* l. At time l a random number ξ_l of new sessions arrive at the buffer. $(\xi_l, l \in \mathbf{Z})$ are independent and identically distributed (i.i.d.) Poisson random variables with mean λ. Session i arriving at time l is active for a random duration $\tau_l(i)$, i.e. it is active during the slots $l, l+1, \ldots, l+\tau_l(i)-1$. Here the $(\tau_l(i), l \in \mathbf{Z}, 1 \leq i \leq \xi_l)$ are i.i.d. random variables taking values in $\mathbf{N}$. Let p_m denote $P(\tau_l(i) = m)$ and $q_m = \sum_{k=m}^{\infty} p_k$. The session duration random variables are assumed to be regularly varying with finite mean and infinite variance, i.e. there is a slowly varying function $L(\cdot)$ and $1 < \alpha < 2$ such that

$$q_k \sim k^{-\alpha}L(k) .$$

(See the appendix for the basic definitions on regular variation.) Each session active during a slot generates traffic at rate 1 during that slot. This completes the description of the arrival model. Note that, since there are many different slowly varying functions, there is considerable flexibility in the choice of arrival model.

We now introduce some auxiliary variables which will facilitate the subsequent discussion.

Let $\xi_l(m)$ denote the number of sessions starting at time l that are active for m slots. We have $\xi_l = \sum_{m=1}^{\infty} \xi_l(m)$. Also $(\xi_l(m), l \in \mathbf{Z}, m \geq 1)$ are seen to be independent Poisson random variables, with $\xi_l(m)$ having mean λp_m. Note that every such session will cease being active at time $l+m$. We define $\eta_l(m)$ as the number of sessions that are active for m slots and cease being active at time l. With this notation, we have $\xi_l(m) = \eta_{l+m}(m)$. We further define $\eta_l = \sum_{m=1}^{\infty} \eta_l(m)$. Note that η_l is also Poisson with mean λ and that $(\eta_l, l \in \mathbf{Z})$ are i.i.d.

Let $\chi_l(m)$ denote the number of sessions starting at time l that are active during slot $l+m-1$. Then $(\chi_l(m), l \in \mathbf{Z}, m \geq 1)$ are Poisson random variables with $\chi_l(m)$ having mean λq_m. Further, we have

$$\chi_l(1) \geq \chi_l(2) \geq \ldots \to 0 \quad a.s.$$

and

$$\chi_l(m) = \sum_{k=m}^{\infty} \xi_l(k) .$$

Note that $\chi_l(1) = \xi_l$.

Let ϕ_l denote the total work brought in by sessions starting at time l. Thus $\phi_l =$

$\sum_{k=1}^{\infty} \chi_l(k) = \sum_{k=1}^{\infty} k\xi_l(k)$. Similarly, let ψ_l denote the total work brought in by sessions that cease to become active at time l. Then $\psi_l = \sum_{k=1}^{\infty} k\eta_l(k)$.

Let $u_l(m)$ denote the total work brought in during slot $l+m-1$ by sessions that started at times prior to l. Then $(u_l(m), l \in \mathbf{Z}, m \geq 1)$ are seen to be Poisson random variables with $u_l(m)$ having mean $\lambda \sum_{k=m+1}^{\infty} q_k$. Further, we have

$$u_l(1) \geq u_l(2) \geq \ldots \to 0 \quad a.s.$$

and

$$u_l(m) = \sum_{n=-\infty}^{l-1} \chi_n(l+m-n) \; .$$

Let U_l denote the total work brought in during all slots following time l by sessions that started at times prior to l. Thus $U_l = \sum_{m=1}^{\infty} u_l(m)$.

We now describe the service offered at the buffer. The buffer is served by a single work-conserving server whose service rate is $s > 0$. If Y_l denotes the total work already present in the system immediately prior to the arrival of the work during slot l, its evolution is described by the equation

$$Y_{l+1} = \max(0, Y_l + \chi_l(1) + u_l(1) - s) \; . \quad (1)$$

Further, we have

$$u_{l+1}(j) = u_l(j+1) + \chi_l(j+1) \quad , \; j = 1, 2, \ldots$$

The evolution of $((Y_l, u_l(1), u_l(2), \ldots),\ l \in \mathbf{Z})$ is therefore described by a Markov process driven by the i.i.d. sequence $((\chi_l(1), \chi_l(2), \ldots),$ $l \in \mathbf{Z})$. Under the stability condition $s > \lambda \sum_{k=1}^{\infty} q_k$ (which is the same as $s > E[\chi_l(1) + u_l(1)]$) this process can be seen to have a unique stationary distribution. (In fact, under the stability condition, the technique of Loynes [9] (see also Baccelli and Bremaud [1, Sec. 2.2]) yields the existence of a unique stationary regime assuming only that $((\chi_l(1),$ $\chi_l(2), \ldots),\ l \in \mathbf{Z})$ is a stationary ergodic sequence.)

3 Statement of results

We will assume that the sessions are served in FIFO order of arrival, with the order of sessions arriving at the same time being determined in a manner independent of the rest of the process. Our goal is to investigate the sojourn time distribution, Σ^*, of the typical session, and more particularly to study the asymptotics of $P(\Sigma^* > x)$ in x. We need to distinguish between two cases : $s \leq 1$ and $s > 1$. If $s \leq 1$, once a session enters service no other session can receive service till its departure. This is not true if $s > 1$.

Case $s \leq 1$:

We first consider the case $s \leq 1$. We remark that the asymptotics of the tail probability of Σ^* can be explicitly determined from classical results in queueing theory.

Let Z_l denote $Y_l + U_l$. This has the interpretation of the remaining work at time l that it is already known will have to be handled by the system. In the case $s \leq 1$ we have

$$Z_{l+1} = \max(0, Z_l + \phi_l - s) \; .$$

Thus the unique stationary regime of $(Z_l, l \in \mathbf{Z})$ is given by

$$Z_l^* = \max(0, \max_{k \geq 1}(\sum_{n=l-k}^{l-1} \phi_n - ks)) \quad (2)$$

(see [1, Sec. 2.2]).

One sees that

$$\phi_l = \sum_{i=1}^{\xi_l} \tau_l(i) \; .$$

Thus we have the following lemma.

Lemma 3.1 *It holds that*

$$P(\phi_l > x) \sim \lambda x^{-\alpha} L(x) \; .$$

Proof:
See [6, Problem 31, pg. 288]. □

Observe now that $(\phi_l, l \in \mathbf{Z})$ are i.i.d. Further, we have $E\phi_l = \lambda E\tau < s$. From Lemma 3.1, ϕ_l has regularly varying tail. Thus we are in the situation addressed by Cohen [3]. We therefore get the following lemma.

Lemma 3.2 *It holds that*

$$P(Z_l^* > x) \sim \frac{\lambda}{(\alpha-1)(s-\lambda E\tau)} x^{1-\alpha} L(x) .$$

Proof:
See [3, Eqn. (3.18)]. □

We may now arrive at the following result.

Theorem 3.1 *If $s \leq 1$, it holds that*

$$P(\Sigma^* > x) \sim \frac{\lambda}{(\alpha-1)s^{\alpha-1}(s-\lambda E\tau)} x^{1-\alpha} L(x) .$$

Proof:

Define σ_l so that $l + \sigma_l$ is the first time at which any session arriving at time l will enter service. With σ_l^* denoting the corresponding stationary quantity, we have $\sigma_l^* = Z_l^*/s$. Now, the total time spent serving all the sessions that arrive at time l is ϕ_l/s. We have

$$\sigma_l^* \leq_{\rm st} \Sigma^* \leq_{\rm st} \sigma_l^* + \frac{\phi_l}{s} .$$

By Lemma 3.1 and Lemma 3.2, ϕ_l has a tail probability that decays strictly faster than that of Z_l^*. Further, ϕ_l is independent of Z_l^*. Thus $P(\Sigma^* > x) \sim P(\sigma_l^* > x)$. From Lemma 3.2 the result follows. □

Case $s > 1$:

We next consider the case $s > 1$. Here we will only derive an asymptotic upper bound for the tail probability $P(\Sigma^* > x)$.
Let Z_l denote $Y_l + U_l$. From equation (1), we have

$$Z_{l+1} = \max(U_{l+1}, Z_l + \phi_l - s) .$$

Thus the unique stationary regime of $(Z_l, l \in \mathbf{Z})$ is given by

$$Z_l^* = \max(U_l, \max_{k \geq 1}(U_{l-k} + \sum_{n=l-k}^{l-1} \phi_n - ks)) .$$

One sees after an algebraic manipulation that

$$\begin{aligned} Z_l^* \leq \quad & \max(0, \max_{k\geq 1}(\sum_{n=l-k+1}^{l} \psi_n - ks)) \\ & + \sum_{j=1}^{\infty} \sum_{m=j+1}^{\infty} m\eta_{l+j}(m) . \qquad (3) \end{aligned}$$

Further, the two terms on the right hand side of equation (3) are independent.
Before stating our upper bound in this case, we will first need the following lemma.

Lemma 3.3 *It holds that*

$$P(\sum_{j=1}^{\infty} \sum_{m=j+1}^{\infty} m\eta_{l+j}(m) > x) \sim \frac{\lambda\alpha}{\alpha-1} x^{-\alpha+1} L(x) .$$

Proof:
Let $\tilde{\eta}$ be Poisson with mean $\lambda(E\tau - 1)$. Let $\tilde{\tau}$ be an $\mathbf{N}$ valued random variable with distribution

$$P(\tilde{\tau} = k) = \frac{(k-1)p_k}{E\tau - 1}$$

and let $\tilde{\tau}(1), \tilde{\tau}(2), \ldots$ be i.i.d. with the distribution of $\tilde{\tau}$ and independent of $\tilde{\eta}$. Then one sees that

$$\sum_{j=1}^{\infty} \sum_{m=j+1}^{\infty} m\eta_{l+j}(m) \stackrel{\rm d}{=} \sum_{i=1}^{\tilde{\eta}} \tilde{\tau}(i) .$$

Further, one has

$$P(\tilde{\tau} > x) \sim \frac{\alpha}{(\alpha-1)(E\tau-1)} x^{-\alpha+1} L(x) .$$

From [6, Problem 31, pg. 288], the claim follows.

□

The following is our result in the case $s > 1$.

Theorem 3.2 *If $s > 1$, it holds that*

$$P(\Sigma^* > x) \lesssim \frac{\lambda(1 + \alpha(s - \lambda E\tau))}{(\alpha - 1)s^{\alpha-1}(s - \lambda E\tau)} x^{1-\alpha} L(x) \ .$$

Proof:

Define σ_l so that $l + \sigma_l$ is the first time at which any session arriving at time l will enter service. With σ_l^* denoting the corresponding stationary quantity, we have $\sigma_l^* \leq Z_l^*/s$. We next remark that once any session arriving in slot l starts receiving service, there is a guaranteed minimum amount of service effort that sessions arriving in slot l will receive in all subsequent slots as long as they have work outstanding. This is because it must be the case that the total number of active sessions among all those that arrived at times strictly before l can now be at most $\lceil s^{-1} \rceil - 1$. Thus there is a fixed constant $\alpha(s)$ (depending only on s) such that all the work in sessions that arrive at time l will depart within a time at most $(\phi_l/\alpha(s)) + 2$ of the time that it started to get served. We have

$$\Sigma^* \leq_{\text{st}} \sigma_l^* + \frac{\phi_l}{\alpha(s)} + 2 \ . \qquad (4)$$

The asymptotics of the tail probability of the first term on the right hand side of equation (3) can be handled by Lemma 3.2, in view of the remark that $(\psi_l, l \in \mathbf{Z})$ has the same distribution as $(\phi_l, l \in \mathbf{Z})$, cf. equation (2). Together with Lemma 3.3, and the independence of the two terms on the right hand side of equation (3), appealing to the proposition in [6, pg. 278] yields

$$P(Z_l^* > x) \lesssim \frac{\lambda(1 + \alpha(s - \lambda E\tau))}{(\alpha - 1)(s - \lambda E\tau)} x^{1-\alpha} L(x) \ . \qquad (5)$$

By Lemma 3.1, ϕ_l has a tail probability that decays strictly faster than the upper bound on the right hand side of equation (5). Further, ϕ_l is independent of Z_l^*. From equation (4) and these observations the result follows.

□

4 Concluding remarks

We have studied the tail of the sojourn time distribution of a typical session in a class of models of a FIFO ATM buffer fed by long-range dependent input traffic. The asymptotics of this tail probability were precisely determined in the regime $s \leq R$, where once a session enters service no work brought in by any other session can receive service before this session completes service. A simple asymptotic upper bound for the tail probability was derived in the regime $s > R$. Studying the regime $s > R$ in more detail is likely to be of interest.

References

[1] Baccelli, F., and Brémaud, P. (1994). *Elements of Queueing Theory : Palm-Martingale Calculus and Stochastic Recurrences.* Springer-Verlag, Berlin.

[2] Beran, J., Sherman, R., Taqqu, M. S., and Willinger, W. (1995). Long-Range Dependence in Variable-Bit-Rate Video Traffic. *IEEE Transactions on Communications*, Vol. 43, No. 2/3/4, pp. 1566 -1579.

[3] Cohen, J. W. (1973). Some Results on Regular Variation for Distributions in Queueing and Fluctuation Theory. *Journal of Applied Probability*, Vol 10, pp. 343 -353.

[4] Cox, D. R. (1984). Long-Range Dependence : A Review. In *Statistics : An Appraisal*, edited by H. A. David and H. T. David, Iowa State University Press, pp. 55 -74.

[5] Duffield, N. G., and O'Connell, N. (1993). Large Deviations and overflow probabilities for the general single-server queue, with applications. To appear in *Proceedings of the Cambridge Philosophical Society*, 1995.

[6] Feller, W. (1971). *An Introduction to Probability Theory and its Application.* John Wiley and Sons, New York.

[7] 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.

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

[9] 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.

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

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

[12] Resnick, S. and Samorodnitsky, G. (1994). The Effect of Long Range Dependence in a Simple Queueing Model. *Preprint*, Cornell University, 14 pp.

Appendix

Regular Variation

For convenience we reproduce the basic definitions on nonnegative random variables with regularly varying tail. For more information, consult Feller [6, pp. 275–284].

Definition 1 *A positive (not necessarily monotone) function $L(\cdot)$ defined on an interval $[a,\infty)$ is said to be* slowly varying *if for every $x > 0$ it holds that*

$$\lim_{t\to\infty} \frac{L(tx)}{L(t)} = 1 \ .$$

Note that there are many slowly varying functions. For instance, any power of any iterated logarithm is slowly varying; any positive function that approaches a strictly positive limit at ∞ is slowly varying.

Definition 2 *A positive function $G(\cdot)$ defined on an interval $[a,\infty)$ is said to be* regularly varying *with exponent $-\alpha$ if it is of the form*

$$G(x) = x^{-\alpha} L(x)$$

where $L(\cdot)$ is a slowly varying function.

In this paper we are only interested in the case $1 < \alpha < 2$.

Definition 3 *A nonnegative random variable X having cumulative distribution function $F(x) = P(X \leq x)$ is said to have* regularly varying tail *if $1 - F(x)$ is regularly varying.*