Ann. Inst. Statist. Math.
Vol. 45, No. 1, 1–7 (1993)

HEAVY AND LIGHT TRAFFIC IN FLUID MODELS WITH BURST ARRIVALS

KARL SIGMAN[1]* AND GENJI YAMAZAKI[2]**

[1] *Department of Industrial Engineering and Operations Research, Columbia University, Mudd Building, New York, NY 10027, U.S.A.*
[2] *Tokyo Metropolitan Institute of Technology, 6-6 Asahigaoka, Hino, Tokyo 191, Japan*

(Received April 11, 1991; revised March 2, 1992)

Abstract. We consider the problem of finding a heavy and light traffic limits for the steady-state workload in a fluid model having a continuous burst arrival process. Such a model is useful for describing (among other things) the packetwise transmission of data in telecommunications, where each packet is approximated to be a continuous flow. Whereas in a queueing model, each arrival epoch, t_n, corresponds to a customer with a service time S_n, the burst model is different: each arrival epoch, t_n, corresponds to a burst of work, that is, a continuous flow of work (fluid, information) to the system at rate 1 during the time interval $[t_n, t_n + S_n]$. In the present paper we show that the burst and queueing models share the same heavy-traffic limit for work, but that their behavior in light traffic is quite different.

Key words and phrases: Fluid model, queue, burst arrivals, heavy traffic, light traffic.

1. Introduction

In Sigman and Yamazaki (1992) a fluid model with a continuous *burst* arrival process was presented and analyzed using sample path techniques to obtain a variety of steady-state type relations concerning work in system. The model was motivated by packetwise transmission of data in telecommunications (where it's applications lie) and was preceded by several related papers (Pan *et al.* (1989), Brandt *et al.* (1990) and Miyazawa and Yamazaki (1992)). In Yamazaki *et al.* (1993), the same model was then analyzed in a stochastic setting (using the Rate Conservation Law) and a number of interesting results were obtained concerning the number of active bursts in the system. In the present paper, we derive some

* Research supported by the Japan Society for the Promotion of Science, during the author's fellowship in Tokyo.

** Research funded by C & C Information Technology Research Laboratories, NEC, and the International Science Foundation.

heavy and light traffic results concerning the steady-state total work in system in the case of renewal input.

Using notation as in Sigman and Yamazaki (1992) (where the reader is referred to for details) we now briefly describe the burst model. The arrival process, $\psi = \{(t_n, S_n) : n \geq 0\}$, is viewed as a marked point process (mpp) with arrival times $0 = t_0 \leq t_1 \leq t_2 \cdots$, marks, $S_n \in \mathcal{R}_+$ and counting process $N(t)$. Each epoch, t_n, starts its own burst, of length S_n, that is, a continuous flow of work to a common processor at rate 1 (during the interval of time $I_n \overset{\text{def}}{=} [t_n, t_n + S_n]$). The processor processes work at rate 1. If at any time t, it holds that $t \in I_n$, then we say that the n-th burst is *active* at time t. The service discipline is FIFO: the server processes all the work from each burst one at a time and does so in the order of the t_n. The burst currently being processed is said to be *in service*. Work that flows in (arrives) from any bursts not in service waits in the queue (of unlimited size). $V_b(t)$ (b for *burst*) denotes the total amount of unprocessed work in the system at time t. We say that the server is *busy* at time t if work is being processed, *idle* otherwise. $V_{b,s}(t)$ denotes the amount of unprocessed work in service at time t and $V_{b,q}(t)$ denotes the amount waiting in queue:

$$V_b(t) = V_{b,q}(t) + V_{b,s}(t).$$

It is important to observe that $V_{b,s}(t) = 0$ if and only if either the server is idle at time t (that is, $V_b(t) = 0$ and there are no active bursts) or the system was idle at epoch t_n- with $t \in [t_n, t_n + \min(S_n, t_{n+1} - t_n)]$. As in Sigman and Yamazaki (1992), we also construct from ψ the corresponding regular FIFO single channel queue (r) and infinite channel queue (∞), with V_r and V_∞ denoting the corresponding workload processes.

In Sigman and Yamazaki (1992) (Proposition 2.2) the following fundamental relationship among the three models is proved:

Proposition 1.1.

$$V_b(t) = V_r(t) - V_\infty(t); \qquad t \geq 0. \tag{1.1}$$

2. Heavy traffic

We shall now show that the burst model work has the same heavy traffic limit as work in the regular queueing model. We proceed with the same kind of heavy-traffic set-up as in Asmussen (1987). Let $\Rightarrow$ denote weak convergence. Consider the $GI/GI/$ burst model in *heavy-traffic*, that is, we assume that we have burst models parameterized by integers $k \geq 1$ with renewal input mpp ψ_k having interarrival times, $T_n(k)$, i.i.d. $\sim A_k \Rightarrow A$ as $k \to \infty$, independent of the marks (burst lengths), $S_n(k)$, i.i.d. $\sim G_k \Rightarrow G$ where A_k has mean $0 < {\lambda_k}^{-1} < \infty$, A has mean $0 < \lambda^{-1} < \infty$, G_k has mean $0 < {\mu_k}^{-1} < \infty$ and G has mean μ^{-1}. ψ denotes the renewal mpp with distributions A, G. We assume $\rho_k \overset{\text{def}}{=} \lambda_k/\mu_k < 1$, and $\lambda/\mu = 1$ and the following heavy traffic condition holds:

$$\lim_{k \to \infty} \rho_k = 1.$$

Letting $S(k)$, $T(k)$, S, T denote generic service and interarrival time r.v.s. from ψ_k and ψ respectively, we also assume a *uniform integrability condition*

$$0 < \lim_{k\to\infty} \operatorname{Var}(T(k) - S(k)) = \operatorname{Var}(T - S) < \infty,$$

so that in particular

$$\gamma_k \overset{\text{def}}{=} \frac{E(T(k) - S(k))}{\operatorname{Var}(T(k) - S(k))} \to 0.$$

In the following, $V_b = V_{b,k}$ denotes a r.v. with the steady-state distribution of work for the burst model with the mpp ψ_k as input (we supress the k for notational simplicity). exp(2) denotes the exponential distribution with mean 1/2.

PROPOSITION 2.1. *As $k\to\infty$,*

$$\gamma_k V_b \Rightarrow \exp(2) \tag{2.1}$$

and

$$E\gamma_k V_b \to \frac{1}{2}. \tag{2.2}$$

PROOF. We take all the mpp's $\psi(k)$, ψ to be time stationary versions on $\mathcal{R}$ on a common probability space. Using (1.1) (via Remark (2.1) in Sigman and Yamazaki (1992)), we then can construct r.v.s. $V_r = V_r(0)$, $V_b = V_b(0)$, $V_\infty = V_\infty(0)$ *all on the same probability space* satisfying

$$V_b = V_r - V_\infty, \tag{2.3}$$

and possessing the steady-state distributions of workloads (we supress the k). It is well known that

$$P(V_r > x) = \rho P(D + S_e > x); \qquad x \geq 0, \tag{2.4}$$

where D and S_e are independent, D has the steady-state customer delay distribution and $S_e \sim G_e$ (the *equilibrium* distribution of service, with density $\mu(1-G(x))$) (see for example, p. 425 in Wolff (1989)). From Asmussen (1987) (Corollary 6.5, p. 199), (together with (2.4)) it follows that (2.1) and (2.2) hold with V_r in place of V_b; thus it suffices to show that the V_∞ term in (2.3) is insignificant in the desired limits. To this end, we use the well known formula

$$E(V_\infty) = \rho E(S^2)/2E(S), \tag{2.5}$$

so that from our uniform integrability condition, we obtain $E\gamma_k V_\infty \to 0$, thus giving (2.2). (2.1) then follows by applying Theorem 4.4.6 of Chung (1974). □

3. Light traffic results

Whereas the heavy traffic limits for b and r are identical, the same is not so in light traffic as we show in this section. We assume the GI/GI set-up with a fixed service time distribution G (with generic service time S), fixed interarrival time distribution A (with generic interarrival time T), and $0 < \rho < 1$. We do *not* assume finite variances. By *light traffic* we mean that the interarrival times, T_n, are scaled by a parameter $\alpha > 0$ to obtain αT_n, where it is assumed that $\alpha \rightarrow \infty$, and hence $\lambda \rightarrow 0$ so that $\rho \rightarrow 0$. In the following, we use the notion $\rho \rightarrow 0$ to mean that $\alpha \rightarrow \infty$.

It is well known (and easy to prove via the sample path methods found on p. 291 Example 5-22 in Wolff (1989)) that $P(V_{r,s} > x) = \rho P(S_e > x)$, where $S_e \sim G_e$ (the *equilibrium* distribution of service, with density $\mu(1 - G(x))$).

LEMMA 3.1. *In light traffic $V_r \approx V_{r,s}$, that is, for all x*

$$\frac{P(V_r > x)}{\rho P(S_e > x)} \rightarrow 1 \quad \textit{as} \quad \rho \rightarrow 0. \tag{3.1}$$

PROOF. From (2.4)

$$\frac{P(V_r > x)}{\rho P(S_e > x)} = \frac{P(D + S_e > x)}{P(S_e > x)}.$$

Since $D \Rightarrow 0$ as $\rho \rightarrow 0$, (3.1) follows. □

PROPOSITION 3.1. *As $\rho \rightarrow 0$,*

$$\frac{P(V_b > x)}{P(V_r > x)} \rightarrow 0. \tag{3.2}$$

PROOF. Let $L = L(0)$, where $L(t)$ denotes a steady-state version for number of customers in the r system and is taken to be constructed on the same probability space as the work processes. Using the fact that $\{V_b > x; L = 1\} = \{V_{b,s} > x; L = 1\}$, we obtain

$$\begin{aligned} P(V_b > x) &= P(V_{b,s} > x; L = 1) + P(V_b > x; L \geq 2) \\ &\leq P(V_{b,s} > x) + P(L \geq 2). \end{aligned} \tag{3.3}$$

From Corollary 2.1 of Sigman and Yamazaki (1992), we have

$$P(V_{b,s} > x) = \rho P(D > x) P(S_e > x). \tag{3.4}$$

Also, it is known that

$$P(L \geq 2) = \rho P(D + S_e > T), \tag{3.5}$$

where T denotes a generic interarrival time and D, S_e, T are independent (see for example, p. 433 of Wolff (1989)). From (3.1)–(3.5) we obtain the following asymptotic upper bound for (3.2):

$$P(D > x) + \frac{P(D + S_e > T)}{P(S_e > x)}, \tag{3.6}$$

which completes the proof since both $P(D > x)$ and $P(D + S_e > T)$ tend to zero in light traffic. □

We next explore a bit deeper into how V_b tends to zero in light traffic.

Proposition 3.2. *If for some critical value $\tilde{\rho}$ it holds that $P(T > S) = 1$; $\rho < \tilde{\rho}$, then in fact*

$$P(V_b = 0) = 1; \qquad \rho < \tilde{\rho}.$$

Proof. If $P(T > S) = 1$ then a.s., $D_n = 0$; $n \geq 0$, where D_n denotes the delay in queue of the n-th customer for r. Every burst thus enters service immediately upon arrival in which case the service rate and the flow rate cancel one another so that a.s., $V_b(t) = 0$; $t \geq 0$. □

Since for r, $V_r \approx V_{r,s}$ in light traffic (via Lemma 3.1), one might expect intuitively the analogous behavior for b, that is, that $V_b \approx V_{b,s}$ so that by (3.4), $P(V_b > x)/(\rho P(D > x)P(S_e > x)) \rightarrow 1$. This is far from being true as we show in

Proposition 3.3. *In the $M/G/$ set-up*

$$\lim_{\rho \rightarrow 0} \frac{P(V_b > 0)}{P(V_{b,s} > 0)} = 1 + \frac{\mu^2}{2} E(S^2), \tag{3.7}$$

so that in particular, if G has infinite second moment then the limit is infinite.

Before proving this proposition, we point out that in real applications, G will indeed have finite second moment so that the above proposition could be used in practice by saying that

For ρ small, $P(V_b > 0)$ is approximately equal to $\rho^2 \left(1 + \frac{\mu^2}{2} E(S^2)\right)$.

Proof. We use Proposition 2.3 in Sigman and Yamazaki (1992) which states (in a more general setting) that

$$P(V_b = 0) = 1 - \rho + \lambda \pi_0 E \min(S, T), \tag{3.8}$$

where $\pi_0 \overset{\text{def}}{=} P(D = 0)$ ($= 1 - \rho$ in the case of Poisson arrivals). Letting $\hat{G}(y) \overset{\text{def}}{=} Ee^{-yS}$, $y > 0$, denote the Laplace transform of G, we obtain $E\min(S,T) = (1 - \hat{G}(\lambda))/\lambda$ so that (3.8) leads to

$$P(V_b > 0) = \rho - (1 - \rho)(1 - \hat{G}(\lambda)).$$

From (3.4) we obtain $P(V_{b,s} > 0) = \rho^2$ and hence

$$\frac{P(V_b > 0)}{P(V_{b,s} > 0)} = \frac{\rho - (1-\rho)(1-\hat{G}(\lambda))}{\rho^2}.$$

Applying L'Hospital's rule twice to evaluate the limit gives (3.7). □

PROPOSITION 3.4. *In the $GI/M/$ set-up,*

$$\overline{\lim_{\rho\to 0}} \frac{P(V_b > x)}{P(V_{b,s} > x)} \le \frac{e^{2\mu x}}{x\mu}, \qquad x > 0. \tag{3.9}$$

In particular, it is always finite.

PROOF. Since for the $GI/M/1$ queue, $(D|D > 0) \sim \exp(\mu\pi_0)$, and $S_e \sim \exp(\mu)$, we obtain for $x > 0$ the following upper bound by using Chebyshev's inequality ($P(V_b > x) \le EV_b/x$) together with the result (proved in a more general setting as Proposition 2.1 in Sigman and Yamazaki (1992)) that $EV_b = \rho ED$:

$$\begin{aligned} P(V_b > x)/P(V_{b,s} > x) &\le \frac{ED}{xP(S_e > x)P(D > x)} \\ &= \frac{e^{2\mu x}}{x\mu\pi_0}. \end{aligned}$$

Since $\pi_0 \to 1$ as $\rho \to 0$, the result follows. □

Remark 3.1. In general, since $V_{b,s} \le V_b$, we can immediately obtain the following lower bound by applying (3.4):

$$\underline{\lim_{\rho\to 0}} \frac{P(V_b > x)}{\rho P(S_e > x)P(D > x)} \ge 1.$$

Remark 3.2. Different light traffic definitions other than scaling interarrival times could be used to obtain similar results to those above. For example, one might consider scaling the service times to zero or thinning the arrival process. The reader is referred to Asmussen (1992) for the most recent approach to light traffic concerning steady-state delay, D, in the $GI/GI/1$ queue.

Remark 3.3. Explicitly evaluating the limit in (3.7) for the $GI/M/$ model does not appear to be as easy as for the $M/G/$ model (nor is our bound in (3.9) of much use); however, we can obtain a reasonable upper bound in the case when A is NWUE (new worse than used in expectation). We sketch the result here where we now scale service times to zero as our light traffic definition (hence $\mu \to \infty$): If A is NWUE, then $\sigma \overset{\text{def}}{=} 1 - \pi_0 \ge \rho$ (see p. 482 of Wolff (1989)). Therefore, since $P(V_b > 0)/P(V_{b,s} > 0)$ can be shown to be a non-increasing function of σ (via (3.8)), it has the upper bound $(1 - a + a\rho)/\rho$, where $a \overset{\text{def}}{=} \mu E\min(S, T) = 1 - \hat{A}(\mu)$.

This simplifies to $1 - \hat{A}(\mu) + (\mu/\lambda)\hat{A}(\mu)$. Assuming that A has a density $f(x)$ (to speed up our derivation, not essential to derive such a bound) we thus obtain as $\mu \to \infty$ the asymptotic upper bound

$$\overline{\lim_{\rho\to 0}} \frac{P(V_b > 0)}{P(V_{b,s} > 0)} \leq 1 + f(0+)/\lambda.$$

We finally point out that if A is NBUE (new better than used in expectation) then a similar argument yields the lower bound

$$\underline{\lim_{\rho\to 0}} \frac{P(V_b > 0)}{P(V_{b,s} > 0)} \geq 1 + f(0+)/\lambda.$$

References

Asmussen, S. (1987). *Applied Probability and Queues*, Wiley, New York.

Asmussen, S. (1992). Light traffic equivalence in single server queues, *Annals of Applied Probability*, **2**, 3, 555–574.

Brandt, A., Brandt, M. and Sulanke, H. (1990). A single server model for packetwise transmission of messages, *Queueing Systems Theory Appl.*, **6**, 287–310.

Chung, K. L. (1974). *A Course in Probability Theory*, 2nd ed., Academic Press, Orlando, California.

Miyazawa, M. and Yamazaki, G. (1992). Loss probability of a burst arrival finite queue with synchronized service, *Probability in the Engineering and Information Sciences*, **6**, 201–216.

Pan, H., Okazaki, H. and Kino, I. (1989). Analysis of bursty traffic in ATM (preprint in Japanese).

Sigman, K. and Yamazaki, G. (1992). Fluid models with burst arrivals: a sample path analysis, *Probability in the Engineering and Information Sciences*, **6**, 17–27.

Wolff, R. W. (1989). *Stochastic Modeling and the Theory of Queues*, Prentice Hall, Englewood Cliffs, New Jersey.

Yamazaki, G., Miyazawa, M. and Sigman, K. (1993). The first few moments of work-load in fluid models with burst arrivals (preprint).