

# Performance Analysis of a Fluid Queue with Random Service Rate in Discrete-Time

Onno J. Boxma

Dept of Mathematics and Computer Science  
and EURANDOM  
Eindhoven University of Technology  
P.O. Box 513, 5600 MB Eindhoven  
The Netherlands.  
Email: boxma@win.tue.nl

Vinod Sharma, D. K. Prasad

Dept Elect. and Comm. Engg.  
Indian Institute of Science  
Bangalore, 560012, India  
Email: {vinod, dkp}@ece.iisc.ernet.in

**Abstract**—We consider a fluid queue in discrete time with random service rate. Such a queue has been used in several recent studies on wireless networks where the packets can be arbitrarily fragmented. We provide conditions on finiteness of moments of stationary delay, its Laplace-Stieltjes transform, various approximations under heavy traffic and asymptotics of its tail distribution. Results are extended to the case where the wireless link can transmit in only a few slots during a frame.

**Keywords:** Fluid queue, discrete-time queue, wireless link, moments of delay, heavy-traffic approximations, tail asymptotics.

## I. INTRODUCTION

We consider a discrete-time queue served by a wireless link. The input to the queue is a stochastic fluid. The link rate of the wireless link changes randomly with time. We will assume that the link rate stays constant during a slot. Such a queue has been used to model wireless systems in several previous studies ([10], [21], [22], [24]). Although in practical wireless systems the input arrives at a wireless link as packets, due to varying link rate, the packets need to be fragmented (for efficient utilization of the link) or several packets may be transmitted within a slot, as the case may be. Thus, the packets lose their identity (from the point of view of service at the queue) and it may be convenient to consider all the contents in a queue as a fluid. As is usually done, we will ignore the overheads due to packet fragmentation.

Discrete-time queues, where the identity of packets is retained (i.e., the packets are not fragmented) have been studied extensively ([7], [9], [29]). The discrete-time queue that we study has two differences from the discrete queues studied in the above literature. The packets in our queues can be fragmented and thus packet boundaries have no relevance, and the number of bits that can be served during a slot is variable (due to wireless link). The discrete-time fluid queue we study has many similarities to the queues usually studied in literature, like the GI/GI/1 queue in continuous time ([2], [8]) and fluid queues in continuous time ([1], [6], [16]), but there are also significant differences. Fortunately, the similarities between our model and continuous and discrete queues studied previously are so strong that we can borrow

significantly from previous studies. Interestingly the similarities are stronger with a continuous time queue than with a discrete queue.

Although, as pointed out above, the queue we study has been considered in wireless literature, it has not been studied extensively. We provide conditions for stability, finiteness of moments of queue length and stationary delay, various approximations under heavy traffic and the exact Laplace-Stieltjes transform (LST) of the delay for our queue. We also provide tail asymptotics when the fluid arriving in a slot has a distribution which is subexponential or has light tails.

We also consider a generalization of the queue which is useful for practical wireless systems. Often, a wireless link is not available at all slots, i.e., the time axis is divided into frames made of several slots and only in some of the slots the link can transmit data ([25]). For example, this happens in uplink and downlink of cellular systems using Time Division Multiple Access (TDMA) and in multihop wireless networks ([15], [20]) due to interference in neighbouring links. The performance of such a link will certainly be different from the queue considered so far. Such a system in wireline context (e.g., when the service rate is fixed) has been studied in [7], [26], [27] (see also the references there-in). We will extend most of our results to this queue.

We are currently working on feed-forward networks of the queues studied in this paper.

The paper is organized as follows. Section II presents the model and the preliminaries. Section III studies stationary delay. It provides the finiteness of moments of stationary delay, its LST, various approximations under heavy traffic and asymptotics of its tail probabilities. Section IV considers the discrete queue embedded in a frame and extends most of the results provided in Section III. Section V provides simulation results to show the closeness of approximations.

## II. THE MODEL AND PRELIMINARIES

We consider a discrete-time queue with infinite buffer. At time  $k$ , the queue length is  $q_k$ , the new arrivals are  $X_k$  and the link (service) rate is  $r_k$  during the time slot  $(k, k + 1)$ . We will denote by  $X$  and  $r$  r.v.s (random variables) with the

distribution of  $X_1$  and  $r_1$  respectively. Then

$$q_{k+1} = (q_k + X_k - r_k)^+ \quad (1)$$

where  $(x)^+$  denotes  $\max(0, x)$ . We will assume  $\{X_k, k \geq 0\}$  and  $\{r_k, k \geq 0\}$  are iid (independent, identically distributed) and independent of each other. However for the stability results in the paper we will only assume that  $\{X_k\}$  and  $\{r_k\}$  are stationary, ergodic sequences.

Sometimes one can assume  $X_k$  and  $r_k$  to be nonnegative integers (bits). But the granularity of  $X_k$  and  $r_k$  can often be fine enough such that these can be taken nonnegative real valued. In the following we will do that.

Equation (1) is the well studied Lindley equation ([2], [8]) and  $\{q_k\}$  in (1) corresponds to the waiting time process in a G/G/1 queue. If  $\mathbb{E}[X] < \mathbb{E}[r] < \infty$  (this assumption will be made in the rest of the paper) and  $\{X_k, r_k\}$  is strictly stationary and ergodic, there is a unique stationary distribution of  $q_k$ . Let  $q$  be a r.v. with this stationary distribution. If the queue starts at  $k = 0$  with any initial distribution,  $q_k$  converges to  $q$  in total variation.

From now on we will make the above mentioned independence assumptions on  $\{X_k\}$  and  $\{r_k\}$ . Then ([14], [28], [30]),  $\mathbb{E}[q^{\alpha-1}] < \infty$  if and only if  $\mathbb{E}[X^\alpha] < \infty$  for  $\alpha \geq 1$  and if  $X$  has finite moment generating function (mgf) in a neighbourhood of 0 then so does  $q$ . Various results on asymptotics of tails of distributions are also available which we will discuss in Section III-D.

The epochs when  $q_k = 0$ , are the regeneration epochs for the process  $\{q_k, k \geq 0\}$ . Let  $\tau$  be a regeneration length (it corresponds to the number of customers served in a busy period in a GI/GI/1 queue). Then ([12], [28], [30])  $\mathbb{E}[\tau^\alpha] < \infty$  for  $\alpha \geq 1$ , if and only if  $\mathbb{E}[X^\alpha] < \infty$ . Also,  $\tau$  has a finite mgf in a neighbourhood of 0 if and only if  $X$  has. This provides rates for the convergence of the distribution of  $q_k$  to that of  $q$ . Combined with results on  $\mathbb{E}[q^\alpha] < \infty$ , one can obtain various functional limit theorems ([28]).

For a queueing system the stationary delay distribution is a key performance measure. We use the results provided above to study the delay distribution in the next section.

### III. DELAY DISTRIBUTION

In this section we study the delay distribution for the system described in Section II. We will study delay for the FCFS (First Come First Served) discipline.

For the fluid  $X_k$  arriving at time  $k$ , the first time its contents are served (i.e., the delay of the first bit) is

$$D_k = \inf\{n : r_k + r_{k+1} + \dots + r_{k+n-1} > q_k\}. \quad (2)$$

The last bit of  $X_k$  will wait for

$$\bar{D}_k = \inf\{n : r_k + r_{k+1} + \dots + r_{k+n-1} \geq q_k + X_k\}. \quad (3)$$

Since the bits served in a slot may belong to packets corresponding to different flows in a wireless system, to ensure QoS (Quality of Service) it is important to study both  $\{D_k\}$

and  $\{\bar{D}_k\}$  and not just the delay of some average (typical) bit in  $\{X_k\}$ . Observe that  $\tau$  is also a regeneration length for  $\{D_k\}$  and  $\{\bar{D}_k\}$ . Thus, if  $\mathbb{E}[X_k] < \mathbb{E}[r_k]$  then  $\mathbb{E}[\tau] < \infty$  and  $\tau$  will also be aperiodic. Hence  $\{D_k\}$  and  $\{\bar{D}_k\}$  have unique stationary distributions. We denote by  $D$  and  $\bar{D}$  r.v.s with the stationary distributions of  $\{D_k\}$  and  $\{\bar{D}_k\}$  respectively. We can define  $D$  and  $\bar{D}$  from (2) and (3) by replacing  $q_k$  and  $q_k + X_k$  by  $q$  and  $q + X$  (because of iid  $\{X_k\}$ , the stationary distribution of queue length seen by arriving bits is the same as that of  $q$ ), where  $q, X$  and  $\{r_k\}$  can be taken independent of each other. For convenience we write

$$D = \inf\{n : r_1 + r_2 + \dots + r_n > q\} \quad (4)$$

and

$$D(t) = \inf\{n : r_1 + r_2 + \dots + r_n > t\} \quad (5)$$

for any  $t \geq 0$ . To avoid trivialities we assume  $P[r > 0] > 0$  and  $P[X > 0] > 0$ .

From (4), if  $\mathbb{E}[q] < \infty$  (then from Proposition 1 in Section III-A below,  $E[D] < \infty$  and  $E[\bar{D}] < \infty$ ),

$$\mathbb{E}[D]\mathbb{E}[r] > \mathbb{E}[q] \quad \text{and} \quad \mathbb{E}[\bar{D}]\mathbb{E}[r] \geq \mathbb{E}[q] + \mathbb{E}[X]. \quad (6)$$

We remark that  $\mathbb{E}[D]$  and  $\mathbb{E}[\bar{D}]$  do not satisfy Little's law. Thus the bounds provided in (6) are useful if  $\mathbb{E}[q]$  is known. One can also obtain bounds on higher moments. For example,

$$\mathbb{E}[D^2] \geq \frac{\mathbb{E}[q^2]}{B_2 \mathbb{E}[r^2]} \quad (7)$$

where  $B_2$  is a known constant. Thus if we know moments of  $q$  or have bounds/approximations for them, we will obtain bounds on moments of  $D$  and  $\bar{D}$ . We will study the accuracy of the bounds in (6) in Section III-C.

If we assume  $r_k$  to be exponentially distributed, then  $\mathbb{E}[q]$  equals the mean delay in the M/GI/1 queue and hence is  $\mathbb{E}[X^2]/(2\mathbb{E}[r](1-\rho))$ , where  $\rho = E[X]/E[r]$ . For this case we will provide an exact expression for  $E[D]$  in Section III-B. Exponentially distributed  $r$  is of particular importance in wireless channels because a Rayleigh distributed channel at low SNR could lead to an exponential  $r$ . Thus we will pay special attention to this case throughout the paper.

For the GI/GI/1 queue several approximations for the mean waiting time are available. For example from [17] we get

$$\mathbb{E}[q] \approx \frac{\rho g \mathbb{E}[X](C_X^2 + C_r^2)}{2(1-\rho)} \quad (8)$$

where

$$\begin{aligned} \rho &= \mathbb{E}[X]/\mathbb{E}[r], \quad C_X^2 = \frac{\text{var}(X)}{(\mathbb{E}[X])^2}, \\ g &= \exp\left[-2 \frac{1-\rho}{3\rho} \frac{(1-C_r^2)^2}{C_r^2 + C_X^2}\right], \quad \text{if } C_r^2 < 1 \\ &= \exp\left[-(1-\rho) \frac{C_r^2 - 1}{C_r^2 + 4C_X^2}\right], \quad \text{if } C_r^2 \geq 1 \end{aligned}$$

and  $C_r^2$  is defined as  $C_X^2$ . This approximation can be used with (6) to obtain approximations/bounds for  $\mathbb{E}[D]$ . For exponential  $r$  this approximation reduces to the exact formula provided

above. Under heavy traffic it is close to the exponential approximation provided at the end of Section III-C. If we add  $\mathbb{E}[X]$  on the right side of (8), we get approximations for  $\mathbb{E}[\bar{D}]$ . In Section V we will provide some simulation results to check the accuracy of these approximations.

The LST of  $q$  is also available if  $X_k$  is of phase type. We will obtain from this the LST of  $D$  and  $\bar{D}$  in Section III-B.

In the rest of the section we study the  $D$  and  $\bar{D}$  in more detail. Section III-A provides conditions for finiteness of moments and mgf. Section III-B provides the LST. Section III-C shows that the bounds in (6) can be tight in heavy traffic and in fact the heavy traffic analysis provides a correction term for these bounds. Section III-D studies the tail asymptotics for the distributions of  $D$  and  $\bar{D}$ .

### A. Finiteness of Moments

In this section we provide conditions for finiteness of moments of  $D$  and  $\bar{D}$ .

**Proposition 1.** If  $\mathbb{E}[\tau^{\alpha+1}] < \infty$  for some  $\alpha \geq 1$  then  $\mathbb{E}[D^\alpha] < \infty$  and  $\mathbb{E}[\bar{D}^\alpha] < \infty$ . Also, if  $\mathbb{E}[\exp(\gamma\tau^\alpha)] < \infty$  for some  $\gamma > 0$  and  $\alpha > 0$  then  $\mathbb{E}[\exp(\gamma'D^\alpha)] < \infty$  and  $\mathbb{E}[\exp(\gamma'\bar{D}^\alpha)] < \infty$  for  $\gamma' < \gamma$ .

**Proof.** Since  $\bar{D} \geq D$ , we show the results for  $\bar{D}$ . Observe that regenerations occur when all the backlog in the queue is cleared. Thus, if  $k = 0$  is a regeneration epoch and  $\tau$  is the next regeneration epoch, then  $\bar{D}_k \leq \tau$  for  $0 \leq k \leq \tau$ . Hence

$$\mathbb{E}[\bar{D}^\alpha] = \frac{1}{\mathbb{E}[\tau]} \mathbb{E}\left[\sum_{k=0}^{\tau-1} (\bar{D}_k)^\alpha\right] \leq \frac{1}{\mathbb{E}[\tau]} \mathbb{E}[\tau^{\alpha+1}]$$

and hence  $\mathbb{E}[\bar{D}^\alpha] < \infty$  if  $\mathbb{E}[\tau^{\alpha+1}] < \infty$ .

Similarly,

$$\begin{aligned} \mathbb{E}[\exp(\gamma'\bar{D}^\alpha)] &= \frac{1}{\mathbb{E}[\tau]} \mathbb{E}\left[\sum_{k=0}^{\tau-1} \exp(\gamma'\bar{D}_k^\alpha)\right] \\ &\leq \frac{1}{\mathbb{E}[\tau]} \mathbb{E}[\tau \exp(\gamma'\tau^\alpha)] \\ &\leq \frac{\mathbb{E}[\tau^p]}{\mathbb{E}[\tau]} \mathbb{E}[\exp(q\gamma'\tau^\alpha)] \end{aligned}$$

for any  $p > 1$ ,  $1/p + 1/q = 1$ . Thus  $\mathbb{E}[\exp(\gamma'\bar{D}^\alpha)] < \infty$  if  $\mathbb{E}[\exp(\gamma\tau^\alpha)] < \infty$  for some  $\gamma > \gamma'$ .  $\square$

Thus we obtain that if  $\mathbb{E}[X^{\alpha+1}] < \infty$  for some  $\alpha > 0$  then  $\mathbb{E}[D^\alpha] < \infty$  and  $\mathbb{E}[\bar{D}^\alpha] < \infty$ . Also,  $D$  and  $\bar{D}$  have finite mgf in a neighborhood of 0 if  $X$  has. The next proposition provides a partial converse.

**Proposition 2.** (i) If  $\mathbb{E}[D^\alpha] < \infty$  for some  $\alpha \geq 1$  then  $\mathbb{E}[q^\alpha] < \infty$ .

(ii) If  $\mathbb{E}[\exp(\gamma D)] < \infty$  for some  $\gamma > 0$  then  $\mathbb{E}[\exp(\gamma'q)] < \infty$  for some  $\gamma' > 0$ .

**Proof.** (i) From [12], Chapter 3, Theorem 8.1,  $\mathbb{E}[D^\alpha(t)]/t^\alpha \rightarrow 1/(\mathbb{E}[r])^\alpha$  for  $\alpha \geq 1$ . Thus for any  $\epsilon > 0$  there is a  $T > 0$  such that for  $t > T$ ,

$\mathbb{E}[D(t)^\alpha] > t^\alpha[1/(\mathbb{E}[r])^\alpha - \epsilon]$ . This implies

$$\begin{aligned} \mathbb{E}[D^\alpha] &= \int_0^T \mathbb{E}[D(t)^\alpha] dP_q(t) + \int_T^\infty \mathbb{E}[D(t)^\alpha] dP_q(t) \\ &\geq \left[\frac{1}{(\mathbb{E}[r])^\alpha} - \epsilon\right] \int_T^\infty t^\alpha dP_q(t) \end{aligned}$$

where  $P_q$  is the distribution function of  $q$ . Thus

$$\begin{aligned} \mathbb{E}[q^\alpha] &= \int_0^T t^\alpha dP_q(t) + \int_T^\infty t^\alpha dP_q(t) \\ &\leq T^\alpha + \int_T^\infty t^\alpha dP_q(t) < \infty. \end{aligned}$$

(ii) Let us assume  $\mathbb{E}[\exp(\gamma D)] < \infty$  for some  $\gamma > 0$ . Let  $N(1) = \max\{n : r_1 + \dots + r_n \leq 1\}$  and let  $\{N_k(1), k \geq 1\}$  be iid with the distribution of  $N(1)$ . Then  $N_1(1) + \dots + N_n(1) \leq_{st} D(n)$ . There exists a  $\gamma_1 > 0$  such that  $\mathbb{E}[\exp(\gamma_1 N(1))] < \infty$  ([12], Chapter 2). Let  $\gamma_2 = \min(\gamma, \gamma_1)$ . Choose a  $\gamma' > 0$  such that  $\mathbb{E}[\exp(\gamma_2 N(1))] > \exp(\gamma')$ . Then,

$$\frac{\mathbb{E}[\exp(\gamma_2 D(t))]}{\exp(\gamma't)} \geq \frac{\mathbb{E}[\exp(\gamma_2 N(1))]^{[t]}}{\exp(\gamma'[t])} \frac{1}{\exp(\gamma'(t - [t]))}$$

where  $[t]$  is the integer part of  $t$ . The right side tends to  $\infty$  as  $t \rightarrow \infty$ . Fix a positive  $M < \infty$ . There is a  $T$  such that for  $t > T$ ,

$$\mathbb{E}[\exp(\gamma_2 D(t))] \geq M \exp(\gamma't).$$

Hence

$$\begin{aligned} \mathbb{E}[\exp(\gamma'q)] &= \int_0^T \exp(\gamma't) dP_q(t) + \int_T^\infty \exp(\gamma't) dP_q(t) \\ &\leq \exp(\gamma'T) + \frac{1}{M} \int_T^\infty \mathbb{E}[\exp(\gamma_2 D(t))] dP_q(t) \\ &\leq \exp(\gamma'T) + \frac{1}{M} \mathbb{E}[\exp(\gamma_2 D)] < \infty. \end{aligned}$$

$\square$

From Propositions 1 and 2, using the previously known results for  $q$  mentioned in Section II, we obtain that  $\mathbb{E}[D^\alpha] < \infty$  ( $\mathbb{E}[\bar{D}^\alpha] < \infty$ ) if and only if  $\mathbb{E}[X^{\alpha+1}] < \infty$  for some  $\alpha \geq 1$ . Also,  $D(\bar{D})$  has an mgf in a neighbourhood of 0 if and only if  $q$  has. Some of these results could be obtained from (28) below under the additional assumption  $\mathbb{E}[r^2] < \infty$ .

### B. LST of the Delay Distribution

Let us first consider the case of  $\exp(\lambda)$  distributed  $r_k$ . Then, according to (4),  $D - 1 =$  number of Poisson( $\lambda$ ) arrivals in  $[0, q]$ . Hence

$$\mathbb{E}[z^{D-1}] = \mathbb{E}[e^{-\lambda(1-z)q}]. \quad (9)$$

Using the well-known Pollaczek-Khintchine formula for the LST of the waiting time distribution in the  $M/G/1$  queue, it then follows, with  $\beta(\cdot)$  denoting the LST of the distribution of  $X$ :

$$\mathbb{E}[z^D] = \frac{(1 - \lambda\mathbb{E}[X])(1 - z)z}{\beta(\lambda(1 - z)) - z}.$$

Next we consider the case when  $X_k$  has a rational LST  $\beta(s) = \beta_1(s)/\beta_2(s)$  where  $\beta_2(s)$  is a polynomial of degree  $m$  and  $\alpha(s)$  is the LST of the distribution of  $r$ . From [8], Section II 5.10,

$$\mathbb{E}[e^{-sq}] = \frac{\beta_2(s)}{\beta_2(0)} \prod_{i=1}^m \frac{\xi_i}{\xi_i - s}, \quad \Re(s) \geq 0,$$

where  $\xi_i, i = 1, \dots, m$  are the zeros of

$$1 - \beta(s)\alpha(-s)$$

in the left half plane. If the  $\xi_i$  are different then we rewrite

$$\mathbb{E}[e^{-sq}] = C_0 + \sum_{i=1}^m \frac{C_i \xi_i}{\xi_i - s}$$

and then

$$P[q > t] = \sum_{i=1}^m C_i e^{\xi_i t}, \quad t > 0,$$

$$P[q = 0] = C_0 = 1 - \sum_{i=1}^m C_i.$$

Next we consider the distributions of  $D$  and  $\bar{D}$ . We have

$$\begin{aligned} P[D > n] &= P[q \geq r_1 + \dots + r_n] \\ &= \sum_{i=1}^m C_i \int_0^\infty e^{\xi_i t} dP[r_1 + \dots + r_n \leq t] \\ &= \sum_{i=1}^m C_i \alpha^n(-\xi_i). \end{aligned} \quad (10)$$

Hence the distribution of  $D$  is a mixture of  $m$  geometric distributions with parameters  $\alpha(-\xi_1), \dots, \alpha(-\xi_m)$ . In particular,

$$\mathbb{E}[D] = \sum_{n=0}^{\infty} P[D > n] = \sum_{i=1}^m C_i \frac{1}{1 - \alpha(-\xi_i)}.$$

If  $r$  is exponential then we can get a more explicit expression for  $\mathbb{E}[D]$ . Indeed, from (10)

$$\begin{aligned} \mathbb{E}[D] &= P[D > 0] + \sum_{n=1}^{\infty} P[D > n] \\ &= 1 + \sum_{n=1}^{\infty} P[q \geq r_1 + \dots + r_n] \\ &= 1 + \int_0^\infty \sum_{n=1}^{\infty} P[r_1 + \dots + r_n \leq t] dP_q(t) \\ &= 1 + \frac{\mathbb{E}[q]}{\mathbb{E}[r]} \end{aligned} \quad (11)$$

where the last equality follows from the fact that  $\sum_{n=1}^{\infty} P[r_1 + \dots + r_n \leq t] = t/\mathbb{E}[r]$ . Of course, (12) could also be obtained by taking the derivative of  $\mathbb{E}[Z^D]$  at  $z = 1$  in (9).

One can similarly obtain the LST of  $\bar{D}$  and  $\mathbb{E}[\bar{D}]$  by replacing  $q$  with  $q + X$ ,  $q$  and  $X$  being independent of each other.

If  $X$  and  $r$  can be taken discrete and  $r \leq \bar{r} < \infty$  then (1)

can be rewritten as

$$q_{k+1} = (q_k + Y_k - \bar{r})^+ \quad (13)$$

where

$$Y_k = X_k + \bar{r} - r_k.$$

Now (13) also occurs for the bulk service queue (with bulk size  $\bar{r}$ ) in discrete-time. Then the moment generating function of the stationary distribution of  $q_k$  is given by ([18])

$$Q(Z) = \frac{\sum_{j=0}^{\bar{r}-1} y_j (Z^{\bar{r}} - Z_j)}{Z^{\bar{r}} - Y(Z)} \quad (14)$$

where  $Z_0, \dots, Z_{\bar{r}-1}$  are the roots of the denominator in (14),  $Y(Z)$  is the moment generating function of  $Y_k$  and  $y_j$  are obtained from  $\sum_{j=0}^{\bar{r}-1} y_j (Z_k^{\bar{r}} - Z_k^j) = 0, k=0, \dots, \bar{r}-1$ . Various computational techniques to obtain  $Z_0, \dots, Z_{\bar{r}-1}$  are available in [18] which then provide the mgf of  $q$ . From this, as above, we can obtain the distributions and moments of  $D$  and  $\bar{D}$ .

### C. Heavy Traffic Approximations

In this section we show that in heavy traffic the bound in (6) will indeed be tight. We also obtain upper bounds on higher moments of  $D$  which are valid in heavy traffic. In fact we first obtain bounds which are valid under congestion for any traffic intensity. Hence our bounds are valid whenever there is congestion in the queue thus covering all the cases where the delay is of real concern.

**Proposition 3.** For  $p \geq 1$ ,

$$\lim_{t \rightarrow \infty} \frac{\mathbb{E}[D^p | q > t]}{\mathbb{E}[q^p | q > t]} \leq \frac{1}{(\mathbb{E}[r])^p}.$$

**Proof.** From Gut [12], Chapter 3, Theorem 8.1,

$$\frac{\mathbb{E}[D(t)^p]}{t^p} \rightarrow \frac{1}{(\mathbb{E}[r])^p}.$$

Thus for any  $\delta > 0$  there exists a  $T$  such that for  $t > T$

$$\mathbb{E}[D(t)^p] \leq \left( \frac{1}{\mathbb{E}[r]^p} + \delta \right) t^p.$$

Since  $q$  is independent of  $\{r_k\}$ , for  $t > T$ ,

$$\begin{aligned} \mathbb{E}[D(q)^p | q > t] &= \frac{1}{P[q > t]} \int_t^\infty \mathbb{E}[D(s)^p | q = s] dP_q(s) \\ &\leq \frac{1}{P[q > t]} \int_t^\infty s^p \left( \frac{1}{\mathbb{E}[r]^p} + \delta \right) dP_q(s) \\ &= \left( \frac{1}{\mathbb{E}[r]^p} + \delta \right) \mathbb{E}[q^p | q > t]. \end{aligned} \quad (15)$$

Taking  $t \rightarrow \infty$  and then  $\delta \downarrow 0$  we obtain the result.  $\square$

From (4) we obtain  $\mathbb{E}[D | q > t] \mathbb{E}[r] \geq \mathbb{E}[q | q > t]$ . Then using Proposition 3

$$\lim_{t \rightarrow \infty} \frac{\mathbb{E}[D | q > t]}{\mathbb{E}[q | q > t]} = 1/\mathbb{E}[r]. \quad (16)$$

In the following we show that under heavy traffic the conditioning on  $\{q > t\}$  in (16) can be removed.

Consider a sequence of queues where the distribution of sequence  $\{r_k\}$  is fixed but the  $n^{\text{th}}$  queue is fed an iid input sequence  $\{X_k^{(n)}, k \geq 0\}$  such that  $X_k^{(n)} \leq_{st} X_k^{(n+1)}$  and  $\rho^{(n)} = \mathbb{E}[X_1^{(n)}]/\mathbb{E}[r_1] \nearrow 1$  as  $n \rightarrow \infty$ . In the following  $q^{(n)}$  and  $D^{(n)}$  denote the stationary queue length and delay in the  $n^{\text{th}}$  queue and then  $n \rightarrow \infty$  provides us a heavy traffic result.

**Proposition 4** Under the above assumptions, for  $p \geq 1$ ,

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[(D^{(n)})^p]}{\mathbb{E}[(q^{(n)})^p]} \leq \frac{1}{(\mathbb{E}[r])^p}. \quad (17)$$

**Proof.** We will use (15) in our proof. Observe that  $T$  in (15) depends only on  $\delta$  and the distribution of  $r_k$  but not on that of  $q^{(n)}$ . Thus, for a given  $\delta > 0$ , choose  $T$  such that (15) holds. Since

$$\begin{aligned} \mathbb{E}[(D^{(n)})^p] &= \mathbb{E}[(D^{(n)})^p \mid q^{(n)} > T] P(q^{(n)} > T) \\ &+ \mathbb{E}[(D^{(n)})^p \mid q^{(n)} \leq T] P(q^{(n)} \leq T) \end{aligned} \quad (18)$$

we consider the two terms on the right side of this equality. As  $\rho \nearrow 1$ ,  $P(q^{(n)} \leq T) \rightarrow 0$  for any fixed  $T$ . Also,

$$\mathbb{E}[(D^{(n)})^p \mid q^{(n)} \leq T] \leq \mathbb{E}[(D^{(n)})^p \mid q^{(n)} = T] < \infty$$

and this upper bound is independent of  $n$ . Thus for any given  $\delta_1 > 0$ , we can choose an  $N_1$  such that

$$\mathbb{E}[(D^{(n)})^p \mid q^{(n)} \leq T] P[q^{(n)} \leq T] \leq \delta_1 \quad (19)$$

for any  $n \geq N_1$ . Then, from (15), (18) and (19) for  $n \geq N_1$

$$\begin{aligned} \mathbb{E}[(D^{(n)})^p] &\leq \\ &\left( \frac{1}{\mathbb{E}[r]^p} + \delta \right) \mathbb{E}[(q^{(n)})^p \mid q^{(n)} > T] P(q^{(n)} > T) + \delta_1. \end{aligned} \quad (20)$$

Next we show that for any  $\delta_2 > 0$  there is an  $N_2$  such that  $|\mathbb{E}[(q^{(n)})^p \mid q^{(n)} > T] P(q^{(n)} > T) - \mathbb{E}[(q^{(n)})^p]| < \delta_2$

for all  $n > N_2$ . But this easily follows from the fact that

$$\begin{aligned} \mathbb{E}[(q^{(n)})^p] &= \mathbb{E}[(q^{(n)})^p \mid q^{(n)} > T] P(q^{(n)} > T) \\ &+ \mathbb{E}[(q^{(n)})^p \mid q^{(n)} \leq T] P(q^{(n)} \leq T) \end{aligned}$$

and  $\lim_{n \rightarrow \infty} \mathbb{E}[(q^{(n)})^p \mid q^{(n)} \leq T] P(q^{(n)} \leq T) = 0$ . Thus we obtain

$$\mathbb{E}[(D^{(n)})^p] \leq \left( \frac{1}{(\mathbb{E}[r])^p} + \delta \right) (\mathbb{E}[(q^{(n)})^p] + \delta_2) + \delta_1$$

for any positive  $\delta, \delta_1, \delta_2$  for all  $n \geq N$ ,  $N$  chosen large enough. Therefore,

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[(D^{(n)})^p]}{\mathbb{E}[(q^{(n)})^p]} \leq \frac{1}{(\mathbb{E}[r])^p}.$$

From (4) and Proposition 4 we obtain

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[D^{(n)}]}{\mathbb{E}[q^{(n)}]} = \frac{1}{\mathbb{E}[r]}$$

and hence  $(\mathbb{E}[D^{(n)}]\mathbb{E}[r] - \mathbb{E}[q^{(n)}])/\mathbb{E}[q^{(n)}] \rightarrow 0$ .

The above results show that under heavy traffic the relative error in approximating  $\sum_{k=1}^D r_k$  with  $q$  goes to zero in mean. However, the error itself does not go to zero. Now we provide some information on it. If  $\mathbb{E}[r^2] < \infty$ , then from Gut [12], Chapter 3, Section 10,

$$\lim_{t \rightarrow \infty} \mathbb{E}[R(t)] = \frac{\mathbb{E}[r^2]}{2\mathbb{E}[r]} \quad (21)$$

if  $r$  is nonarithmetic. If  $r$  has an arithmetic distribution on a lattice with span  $d$  then

$$\lim_{k \rightarrow \infty} \mathbb{E}[R(kd)] = \frac{\mathbb{E}[r^2]}{2\mathbb{E}[r]} + \frac{d}{2}.$$

In the following we provide the results for only the nonarithmetic case. For the arithmetic, just add  $d/2$ . Using the above techniques, one can easily show from (21)

$$\lim_{T \rightarrow \infty} \mathbb{E}[R(q) \mid q > T] = \frac{\mathbb{E}[r^2]}{2\mathbb{E}[r]}. \quad (22)$$

The next proposition removes the conditioning in heavy traffic.

**Proposition 5.** If  $\mathbb{E}[r^2] < \infty$ ,

$$\lim_{n \rightarrow \infty} \mathbb{E}[R^{(n)}] = \frac{\mathbb{E}[r^2]}{2\mathbb{E}[r]}.$$

**Proof.** We have

$$\begin{aligned} \mathbb{E}[R^{(n)}] &= \mathbb{E}[R^{(n)} \mid q^{(n)} > T] P(q^{(n)} > T) \\ &+ \mathbb{E}[R^{(n)} \mid q^{(n)} \leq T] P(q^{(n)} \leq T). \end{aligned} \quad (23)$$

Given  $\delta > 0$  from (22) there exists a  $T$  s.t.

$$\left| \mathbb{E}[R^{(n)} \mid q^{(n)} > t] - \frac{\mathbb{E}[r^2]}{2\mathbb{E}[r]} \right| < \delta \quad (24)$$

for all  $t > T$ . This  $T$  depends only on the distribution of  $r$  and  $\delta$  but is independent of  $n$ . We also have  $P(q^{(n)} > T) \nearrow 1$  as  $n \rightarrow \infty$  for any fixed  $T$ .

Next consider the second term on the right side in (23). For any fixed  $T$ ,

$$\mathbb{E}[R^{(n)} \mid q^{(n)} \leq T] \leq \max_{0 \leq t \leq T} \mathbb{E}[R(t)] \quad (25)$$

and  $R(t) \leq r_{D(t)}$ . From [12], Chapter 3, Theorem 7.2,  $\{r_{D(t)}/t, t \geq 1\}$  is uniformly integrable. In particular, the right side of (25) is upper bounded by  $MT$  for some  $M < \infty$ . Thus

$$\left| \mathbb{E}[R^{(n)}] - \frac{\mathbb{E}[r^2]}{2\mathbb{E}[r]} \right| < 2\delta$$

for all  $n$  large enough.  $\square$

Thus  $(\mathbb{E}[q] + \mathbb{E}[r^2]/2\mathbb{E}[r])/\mathbb{E}[r]$  provides a better approximation of  $\mathbb{E}[D]$  under heavy traffic. For exponential  $r$ , this reduces to the exact formula (12).

One can similarly obtain results for higher moments of  $R^{(n)}$ . We also know that  $R(t) \xrightarrow{W} Y$  where  $P[Y \leq x] =$

$\int_0^x (1 - F_r(s)) ds / \mathbb{E}[r]$  and  $\rightarrow^W$  denotes weak convergence. This then will provide us with  $R^{(n)} \rightarrow^W Y$ .

We can also exploit the standard heavy traffic approximations on the GI/GI/1 queue. For example, ([2], Chapter X) we know that if  $\mathbb{E}[r] - \mathbb{E}[X^{(n)}] \rightarrow \nu$  and  $Var(X^{(n)}) + Var(r) \rightarrow \sigma^2$  then

$$2\nu q^{(n)} / \sigma^2 \rightarrow^W \exp(1). \quad (26)$$

Since  $t \rightarrow D(t)$  is a continuous function, this implies that  $D(2\nu q^{(n)} / \sigma^2) \rightarrow^W D(Y)$  where  $Y$  is exponentially distributed with mean 1. In particular, then from (10) in heavy traffic (but with  $\rho < 1$ ),

$$\begin{aligned} P[D > m] &= \int_0^\infty P[D(t) > m] dP_q(t) \\ &= \frac{2\nu}{\sigma^2} \int_0^\infty P[D(t) > m] \exp[-2\nu t / \sigma^2] dt \\ &= \frac{2\nu}{\sigma^2} \int_0^\infty P[r_1 + \dots + r_m \leq t] \exp[-2\nu t / \sigma^2] dt \\ &= \alpha^m (2\nu / \sigma^2) \end{aligned}$$

where  $\alpha(s)$  is the LST of  $r$ . Thus as a heavy traffic approximation,  $D$  is geometrically distributed with parameter  $\alpha(2\nu / \sigma^2)$ . This provides

$$\mathbb{E}[D] = 1 / (1 - \alpha(2\nu / \sigma^2)). \quad (27)$$

Comparing these results with those in Section III-B, one observes that these results are simpler and hold under general assumptions on the distribution of  $X$  but of course provide good approximations only under heavy traffic (interestingly we will see in Section V via simulations that although the approximation for  $\mathbb{E}[q]$  is not good under light traffic, for  $\mathbb{E}[D]$  it is).

Finally consider  $\mathbb{E}[D^2]$ . From (7) and (17)

$$\frac{1}{(\mathbb{E}[r])^2} \geq \frac{\mathbb{E}[D^2]}{\mathbb{E}[q^2]} \geq \frac{1}{B_2 \mathbb{E}[r^2]}$$

under heavy traffic whereas the above exponential approximation provides  $\mathbb{E}[q^2] = \sigma^4 / 2\nu^2$ .

#### D. Asymptotics for Tail Probabilities

In this section we consider the asymptotics on the tail of the distributions of  $D$  and  $\bar{D}$ . In Section III-A we have provided conditions for finiteness of  $\mathbb{E}[D^\alpha]$ ,  $\mathbb{E}[\bar{D}^\alpha]$  and of their mgf in a neighborhood of 0. These provide asymptotics for the tail distributions of  $D$  and  $\bar{D}$  via the relationship: for  $\alpha \geq 1$ ,  $\mathbb{E}[D^\alpha] = \alpha \int_0^\infty x^{\alpha-1} P[D \geq x] dx < \infty$  if and only if  $\int_0^\infty x^\alpha P[X \geq x] dx < \infty$ . Thus  $D$  is of regular variation with index  $-\alpha - 1$  if and only if  $X$  is of regular variation with index  $-\alpha$ . Also,  $D$  has an mgf in a neighbourhood of 0 if  $X$  has. Thus  $D$  has a light tailed (exponentially decaying) distribution if  $X$  has.

Some more information on the tail distribution can be obtained for subexponential distributions. From the results on the waiting time distribution for a GI/GI/1 queue we know

([11], [30]):

If  $X_e$  is subexponential, then

$$P[q > x] \sim \frac{\rho}{1 - \rho} P[X_e > x]$$

where  $X_e$  is a r.v. with the equilibrium distribution of  $X$ :

$$P[X_e > x] = \frac{1}{\mathbb{E}[X]} \int_x^\infty P(X > u) du,$$

and  $f(x) \sim g(x)$  denotes  $\lim_{x \rightarrow \infty} f(x)/g(x) = 1$ .

The asymptotics of  $D$  are more complicated. From (4) we observe that  $D = N(q) + 1$  where  $N(t)$  is the undelayed renewal process formed from  $\{r_k\}$ . Then using results from [3] we obtain:

- 1) From Theorems 3.6, 3.11 and Proposition 3.1 of [3]: If  $\mathbb{E}[r_1^2] < \infty$  and  $q$  has tail heavier than  $e^{-\sqrt{x}}$ , i.e.,  $\lim_{x \rightarrow \infty} P[q > x] e^{+\sqrt{x}} = \infty$  then

$$P[D > x] \sim P\left[q > \frac{x}{\mathbb{E}[r]}\right] \quad \text{as } x \rightarrow \infty. \quad (28)$$

We have observed above that, if  $X_e$  is subexponential then  $q$  has asymptotics of  $X_e$ . Thus if  $X_e$  has tail heavier than  $e^{-\sqrt{x}}$ , (28) holds. Also then  $X_e$  and hence  $q$  has heavier tails than that of  $X$ . Therefore, if  $X$  is also subexponential then  $q$  and  $X$  being independent in defining  $\bar{D}$ ,  $q + X$  is subexponential and has the same tail behaviour as that of  $q$ . Thus  $D$  and  $\bar{D}$  also have the same tail behaviour.

- 2) For  $r_k$  exponentially distributed, there are results in [3] corresponding to  $q$  with heavy tails but lighter than  $e^{-\sqrt{x}}$ . These directly provide corresponding results for  $D$  and  $\bar{D}$ .

#### IV. DISCRETE QUEUE EMBEDDED IN A FRAME

In a practical wireless system it will often happen that a wireless link gets the opportunity to transmit data from its queue only in some of the slots and often these slots appear periodically. For example, this can happen if several queues share a common wireless link in TDMA fashion (e.g., GSM cellular system, the subscriber stations in a WiMAX uplink ([23]) and a multihop wireless network ([15])). These slots can also appear randomly: if the wireless channel is bad in a slot it may be decided not to transmit in that slot in order to save transmit power. In this section we extend the results obtained so far to this setup when a queue is served at periodic intervals.

We assume a frame is made up of  $T$  slots. The queue under consideration gets to serve in the first  $L \leq T$  slots of each frame. In the last  $T - L$  slots the fluid can arrive at the queue but it will need to wait till the next frame for transmission. Let  $q_k$  be the queue length at the beginning of the  $k^{\text{th}}$  frame. Let  $X_{k,i}$  be the new arrivals to the queue in the beginning of the  $i^{\text{th}}$  slot of the  $k^{\text{th}}$  frame and let  $r_{k,i}$  be the link rate in

that slot,  $i = 1, \dots, T$ . Then

$$\begin{aligned} q_k &= q_{k,1}, \\ q_{k,i+1} &= (q_{k,i} + X_{k,i} - r_{k,i})^+, \quad i = 1, \dots, L, \\ q_{k+1} &= (q_{k,L+1} + X_{k,L+1} + \dots + X_{k,T}). \end{aligned} \quad (29)$$

We denote  $Y_k = X_{k,L+1} + \dots + X_{k,T}$ . We assume  $X_{k,i}$  and  $Y_{k,i}$  to be iid for each  $k$  and  $i$  (for the stability result we need to assume them to be only stationary, ergodic, sequences). We will also denote

$$\begin{aligned} X_k &= (X_{k,1}, X_{k,2}, \dots, X_{k,T}), \\ r_k &= (r_{k,1}, r_{k,2}, \dots, r_{k,T}). \end{aligned}$$

Then

$$q_{k+1} = f(q_k, (X_k, r_k))$$

where  $f$  can be found from (29). One can easily see from (29) that  $f$  is nondecreasing and continuous in  $q$ . Thus, from Loynes [19], if  $\{(X_k, r_k)\}$  forms a stationary, ergodic sequence then  $q_k$  has a stationary distribution (which may be infinite with positive probability). Let  $q$  be a r.v. with this stationary distribution. Since sequence  $\{q_k\}$  can be lower bounded by sequence

$$\begin{aligned} \bar{q}_{k+1} &= \left( \bar{q}_k + \sum_{i=1}^T X_{k,i} - \sum_{i=1}^L r_{k,i} \right)^+, \\ \bar{q}_0 &= q_0 \end{aligned} \quad (30)$$

and  $\bar{q}_k$  converges a.s. to  $\infty$  if  $T\mathbb{E}[X_{k,1}] > L\mathbb{E}[r_{k,1}]$ ,  $q_k$  also converges to  $\infty$  a.s. under these conditions.

Next we show that when  $T\mathbb{E}[X_{k,1}] < L\mathbb{E}[r_{k,1}]$ ,  $q$  has a proper distribution, i.e.,  $P(q < \infty) = 1$ . Suppose it is not true. Using Loynes' construction ([19]), we show it leads to a contradiction. Let the  $0^{th}$  frame start at  $k = 0$ . Let  $q_{0,i}^{-k}$  be the queue length at the beginning of the  $i^{th}$  slot in the  $0^{th}$  frame when the system started empty at the beginning of  $-k^{th}$  frame. Then  $q_{0,1}^{-k}$  monotonically increases a.s. and  $q_{0,1}^{-k}(\omega) \nearrow \infty$  on a set  $A$  with  $P(A) > 0$  under our assumptions. Also,  $q_{0,i}^{-k}(\omega) \nearrow \infty$  for  $\omega \in A$  for  $i = 2, \dots, L$ . Thus there is an  $M$  s.t. from  $k > M$ ,  $q_{0,i}^{-k}(\omega)$  will not be zero for  $i = 1, \dots, L$ . Thus from  $k > M$ ,  $q_{0,1}^{-k}(\omega)$  will follow (30) without the superscript  $+$ . But by the Strong Law of Large numbers  $q_{0,1}^{-k}$  will converge to 0 a.s. and hence  $P(A) = 0$ . Thus  $q$  has a proper stationary distribution.

Using the above construction we can also show that when  $T\mathbb{E}[X_{k,1}] < L\mathbb{E}[r_{k,1}]$  the stationary distribution is unique and starting from any initial distribution,  $q_k$  converges in total variation to it. From now on in this section we will assume  $T\mathbb{E}[X_{k,1}] < L\mathbb{E}[r_{k,1}]$ .

From now onwards we will make the independence assumptions on  $\{(X_k, r_k)\}$ . We can rewrite (29) slotwise as in (1) with  $r_{k,i} = 0$  for  $i = L+1, \dots, T$ . Now however the rate sequence is no longer iid but periodic and the results obtained in the previous section cannot be directly used. However, this can be taken as a regenerative sequence with

regeneration epochs the frame boundaries. Then we are within the framework of [28]. Thus we obtain  $\mathbb{E}[\tau^\alpha] < \infty$  ( $\tau$  has a finite mgf in a neighborhood of 0), if  $\mathbb{E}[X_1^\alpha] < \infty$  ( $X$  has a finite mgf in a neighborhood of 0) for  $\alpha \geq 1$  where  $\tau$  is a regeneration length for this system, the regeneration epochs being the frame boundaries where  $q_k = 0$ . We also obtain  $\mathbb{E}[q^\alpha] < \infty$  whenever  $\mathbb{E}[X_1^{\alpha+1}] < \infty$  for  $\alpha \geq 1$  where  $q$  is the stationary queue length at frame boundaries. Finiteness of the mgf of  $q$  in a neighborhood of 0 is also implied by that of  $X_1$ .

Various functional limit theorems for the process  $\{q_k\}$  and rates of convergence to the stationary distribution are also obtained ([28]). Although [28] provides these results for the process observed at slot boundaries, these results at frame boundaries can be obtained easily in the same way. The stationary distribution of the queue length process is different at the  $i^{th}$  slot than at the  $j^{th}$  slot,  $j \neq i$  within a frame but one can easily relate these stationary distributions. Also finiteness of moments of their stationary distributions holds under the same conditions. For example, for  $1 < i \leq L$  (denoting by  $q_i$  a r.v. with the stationary distribution of  $q_{k,i}$ )

$$q_i = ((q + X_1 - r_1)^+ + \dots + X_{i-1} - r_{i-1})^+ \leq q + \sum_{j=1}^{i-1} X_j$$

and hence  $\mathbb{E}[q_i^\alpha] < \infty$  if  $\mathbb{E}[q^\alpha] < \infty$  and  $\mathbb{E}[X^\alpha] < \infty$ . Similarly we obtain the finiteness of the exponential moments.

Let  $D_i$  and  $\bar{D}_i$  be r.v.s obtained from (2) and (3) by replacing  $q_k$  by  $q_i$ . Thus the Propositions 1-5 and the asymptotics on the tail distributions hold as for the system in Section II. Let  $\hat{D}_i$  and  $\hat{\hat{D}}_i$  be the delay of the first bit of  $X_{k,i}$  under stationarity. Let  $D, \bar{D}, \hat{D}, \hat{\hat{D}}$  denote these quantities when  $i = 1$ . Then,

$$\left(\frac{D}{L} - 1\right)T \leq \hat{D} = \left[\frac{D}{L}\right]T + \left(D - \left[\frac{D}{L}\right]L\right) \leq D \frac{T}{L} \quad (31)$$

where  $[x]$  denotes the largest integer  $< x$ . Similarly one can relate  $\hat{D}_i$  and  $D_i$ . Thus we obtain the finiteness of moments of  $\hat{D}_i$  and  $\bar{D}_i$  from that of  $D_i$  and  $\bar{D}_i$ , for which we obtain these results as in Proposition 1 from that of  $q_i$ . Also the distribution of  $\hat{D}$  is of regular variation of index  $-\alpha$  if and only if that of  $D$  is. Similarly  $\hat{D}$  has a mgf in a neighbourhood of 0 if and only if  $D$  has. If we know the distribution of  $D$  then we can use the equality in (31) to obtain the distribution of  $\hat{D}$ . If we have only moments and/or bounds or approximations on moments of  $D$  then the inequalities in (31) can be used to obtain the corresponding bounds/approximations for  $\hat{D}$ .

Under heavy traffic, one expects that the queue will not be empty most of the time. Then, one can approximate (29) by removing  $()^+$  on the RHS of (29). Now observe this queue only at frame boundaries. Consider the frame as a slot and then use the results of Section III with  $X_k = X_{k,1} + \dots + X_{k,T}$  and  $r_k = r_{k,1} + \dots + r_{k,L}$ .

Finally we obtain the LST of  $q$  and  $D$  when  $r$  is exponentially distributed. As commented before, this could correspond to Rayleigh fading channels and hence is of practical concern.

The corresponding results for  $q_i$ ,  $D_i$  and  $\bar{D}_i$  can then be easily obtained by relating them to  $q$ ,  $D$  and  $\bar{D}$ .

Let  $\beta(s)$  be the LST of  $X$ . Then

$$\begin{aligned} & \mathbb{E}[e^{-sq_{k,2}} \mid q_{k,1} = y] \\ &= \mathbb{E}[e^{-s(q_{k,1} + X_{k,1} - r_{k,1})^+} \mid q_{k,1} = y] \\ &= \mathbb{E}[e^{-s(y+X-r)}] - \mathbb{E}[e^{-s(y+X-r)} \mathbf{1}_{\{y+X-r \leq 0\}}] \\ &+ P[y+X-r \leq 0] \\ &= e^{-sy} \beta(s) \frac{\lambda}{\lambda-s} - \frac{\lambda}{\lambda-s} P[y+X-r \leq 0] \\ &+ P[y+X-r \leq 0] \\ &= e^{-sy} \beta(s) \frac{\lambda}{\lambda-s} - \frac{s}{\lambda-s} \beta(\lambda) e^{-\lambda y}. \end{aligned}$$

Thus

$$\mathbb{E}[e^{-sq_{k,2}}] = \beta(s) \frac{\lambda}{\lambda-s} \mathbb{E}[e^{-sq_{k,1}}] - \beta(\lambda) \frac{s}{\lambda-s} \mathbb{E}[e^{-\lambda q_{k,1}}].$$

We can iterate this equation to obtain

$$\mathbb{E}[e^{-sq_{k,L+1}}] = h^L(s) \mathbb{E}[e^{-sq_{k,1}}] - g(s) \sum_{j=0}^{L-1} h^j(s) \mathbb{E}[e^{-\lambda q_{k,L-j}}]$$

where  $h(s) = \beta(s)\lambda/(\lambda-s)$ ,  $g(s) = \beta(\lambda)s/(\lambda-s)$ . Furthermore, we obtain  $\mathbb{E}[e^{-sq_{k,T+1}}] = \beta^{T-L}(s)\mathbb{E}[e^{-sq_{k,L+1}}]$ . In steady state  $\mathbb{E}[e^{-sq_{k,T+1}}] = \mathbb{E}[e^{-sq_{k,1}}] \triangleq \mathbb{E}[e^{-sq}]$ . Thus we obtain

$$\begin{aligned} \mathbb{E}[e^{-sq}] &= -\left[1 - \beta^T(s) \left(\frac{\lambda}{\lambda-s}\right)^L\right]^{-1} \\ &\left\{ \beta^{T-L}(s) \beta(\lambda) \frac{s}{\lambda-s} \sum_{j=0}^{L-1} \left(\frac{\lambda}{\lambda-s}\right)^j y_{L-j}(\lambda) \right\} \end{aligned} \quad (32)$$

The right hand side contains  $L$  unknowns  $y_{L-j}$ ,  $j = 0, \dots, L-1$ . From Rouché's theorem ([8], p 652), we know that under stability,

$$1 - \beta^T(s) \left(\frac{\lambda}{\lambda-s}\right)^L$$

has exactly  $L$  zeros  $s_0, \dots, s_{L-1}$  in the right half plane. The term in curly brackets in the right side of (32) should also be zero for  $s = s_0, \dots, s_{L-1}$ . This yields  $L$  equations that can be used to obtain  $y_1(\lambda), \dots, y_L(\lambda)$ . There are several algorithms available to compute the zeros  $s_0, \dots, s_{L-1}$  (see [18]). It follows, after a lengthy calculation, that

$$\begin{aligned} \mathbb{E}[q] &= (T-L)\mathbb{E}[X] + \frac{1}{L-T\lambda\mathbb{E}[X]} \\ &+ \left[ \beta(\lambda) \sum_{j=0}^{L-1} y_{L-j}(\lambda) \left( j\mathbb{E}[X] + \frac{L-j-1}{\lambda} \right) \right. \\ &\left. + \lambda T(T-1) \frac{\mathbb{E}[X]^2}{2} + \lambda T \frac{\mathbb{E}[X^2]}{2} - \frac{L(L-1)}{2\lambda} \right]. \end{aligned} \quad (33)$$

It should be noted that this formula reduces to the familiar expression for the mean waiting time in the  $M/G/1$  queue when  $T = L = 1$ .

The mgf of  $D$  follows from the LST of  $q$  in (32), using (9):

$$\mathbb{E}[z^D] = \frac{\beta^{T-L}(\lambda(1-z))\beta(\lambda)(1-z) \sum_{j=0}^{L-1} \beta(\lambda(1-z))^j z^{L-j} y_{L-j}(\lambda)}{\beta^T(\lambda(1-z)) - z^L}.$$

If we are able to invert this mgf to obtain the distribution of  $D$  (there are various techniques to perform such a numerical inversion) then the mgf of  $\hat{D}$  follows from (31):

$$\mathbb{E}[z^{\hat{D}}] = \sum_{j=0}^{\infty} \sum_{k=0}^{L-1} z^{jT+k} P(D = jL+k).$$

In fact one can show that

$$\mathbb{E}[z^{L\hat{D}}] = \frac{1}{L} \sum_{k=0}^{L-1} \mathbb{E}[z^{TD} a^k] \frac{1 - (a^{-k} z^{L-T})^L}{1 - a^{-k} z^{L-T}},$$

where  $a = \exp(\frac{2\pi i}{L})$ ,  $i = \sqrt{-1}$ . Thus one can avoid inverting the mgf of  $D$ . One can then obtain

$$\mathbb{E}[\hat{D}] = \frac{1}{L} \frac{d}{dz} \mathbb{E}[z^{L\hat{D}}] \Big|_{z=1}. \quad (34)$$

## V. SIMULATION RESULTS

In this section we provide some simulation results to verify the accuracy of approximations provided in Sections III and IV.

We first consider the single queue studied in Sections II and III. For  $X$  we have taken a few discrete distributions: Poisson (examples 9, 10) and finite valued while for  $r$  we have taken Rayleigh (examples 9, 10), exponential (examples 7, 8) and a few discrete finite valued distributions. The  $\rho$  has been taken as 0.3, 0.5, 0.7, 0.9, 0.95 and 0.98. Each simulation was done for 20 million slots, long enough to have negligible estimation error. The simulated values of  $\mathbb{E}[q]$ ,  $\mathbb{E}[D]$ ,  $\mathbb{E}[q]$  from formula (8) and its heavy traffic approximation  $= (\text{var}(X) + \text{var}(r))/(2(\mathbb{E}[r] - \mathbb{E}[X]))$  and  $\mathbb{E}[D]$  via the lower bound  $\mathbb{E}[q]/\mathbb{E}[r]$  and with the heavy traffic correction  $\mathbb{E}[q]/\mathbb{E}[r] + \mathbb{E}[r^2]/(2(\mathbb{E}[r])^2)$  (with  $d/(2\mathbb{E}[r])$  added if  $r$  has an arithmetic distribution) and as  $1/(1 - \alpha(2\nu/\sigma^2))$  are provided in Tables I and II. As commented before, for  $r$  exponential, (8) and  $\mathbb{E}[D]$  with heavy traffic correction are in fact exact formulae for all traffic intensities. This is also seen from Table II. For other distributions, the theoretical formulae for  $\mathbb{E}[q]$  and  $\mathbb{E}[D]$  are very close to the simulated values for  $\rho \geq 0.90$ . For lower values of  $\rho$  the approximations are not good as expected (but the approximation (27) for  $\mathbb{E}[D]$  is still working quite well). We also observe that  $\mathbb{E}[q]$  from (8) is quite close to simulations even for small  $\rho$  and is always more accurate than the heavy traffic approximation.

Next we provide an example for the queue in Section V. We consider exponential service and arrival rates. We fix the service rate at  $\mathbb{E}[r] = 7.5$  and take different values for  $\mathbb{E}[X]$ . Also, we take  $T = 10$  and  $L = 4$ . We obtained  $\mathbb{E}[q]$  and  $\mathbb{E}[\hat{D}]$

TABLE I  
SIMULATION RESULTS FOR SINGLE QUEUE OF SECTIONS II AND III:  
MEAN QUEUE LENGTH

ex am pl es	$\rho$	EX	Er	var(X)	var(r)	Eq		
						simulated value	From (8)	From heavy traffic appro (26)
1	0.30	0.36	12.00	17.09	21.00	0.61	0.52	2.27
2	0.50	5.00	10.00	20.10	33.75	2.65	2.21	5.38
3	0.70	7.13	10.18	37.36	23.39	6.45	6.69	9.95
4	0.90	7.13	7.92	37.36	32.33	40.23	39.58	44.01
5	0.95	7.13	7.50	37.36	18.75	70.39	71.33	74.81
6	0.98	7.35	7.50	12.23	18.75	100.08	99.70	103.26
7	0.95	7.13	7.50	37.36	56.25	118.00	117.50	124.81
8	0.98	7.35	7.50	12.23	66.25	220.99	220.83	228.26
9	0.95	7.13	7.50	7.13	15.41	27.10	26.81	30.04
10	0.98	7.35	7.50	7.35	15.41	73.48	72.54	75.85

TABLE II  
SIMULATION RESULTS FOR SINGLE QUEUE OF SECTIONS II AND III:  
MEAN DELAY

ex am pl es	$\rho$	EX	Er	var(X)	var(r)	ED			
						simulated value	Eq/Er	with heavy traffic correction	from (27)
1	0.30	0.36	12.00	17.09	21.00	1.12	0.04	0.82	1.12
2	0.50	5.00	10.00	20.10	33.75	1.39	0.22	1.14	1.44
3	0.70	7.13	10.18	37.36	23.39	1.47	0.66	1.32	1.68
4	0.90	7.13	7.92	37.36	32.33	6.01	5.00	6.51	6.35
5	0.95	7.13	7.50	37.36	18.75	10.40	9.51	10.84	10.66
6	0.98	7.35	7.50	12.23	18.75	14.13	13.29	14.63	14.45
7	0.95	7.13	7.50	37.36	56.25	16.74	15.67	16.67	17.64
8	0.98	7.35	7.50	12.23	66.25	30.46	29.44	30.44	31.43
9	0.95	7.13	7.50	7.13	15.41	4.30	3.58	4.21	4.66
10	0.98	7.35	7.50	7.35	15.41	10.46	9.67	10.31	10.76

via simulations and compared with the theory. The theoretical results are obtained via (33) and (34). We also obtained  $\mathbb{E}[q]$  via heavy traffic approximation mentioned Section IV.

TABLE III  
EXAMPLE FOR QUEUE EMBEDDED IN A FRAME

$\rho$	$\mathbb{E}[X]$	$\mathbb{E}[r]$	Simu- lated $\mathbb{E}[q]$	$\mathbb{E}[q]$ from (33)	$\mathbb{E}[q]$ via HT	E $\hat{D}$			
						Simu- lated	via (34)	via (33) $\hat{D}_{upper}$	via HT $\hat{D}_{upper}$
0.80	2.40	7.50	28.45	28.44	23.55	9.00	9.60	10.35	10.35
0.90	2.70	7.50	55.06	55.34	49.65	17.48	18.63	18.45	19.05
0.95	2.85	7.50	107.40	108.15	102.17	34.74	36.27	36.05	36.56
0.98	2.94	7.50	262.17	265.83	259.53	86.21	88.85	88.61	89.01

The results of the simulations and theory are presented in Table III. The simulated values  $\mathbb{E}[\hat{D}]$  are closer to the upper bounds than to the lower bounds for  $\mathbb{E}[\hat{D}]$  when we used (33) and hence we present only the upper bounds. We further observe that the heavy traffic approximations are also providing good approximations.

**Acknowledgement** The work of the second author was done while he was visiting EURANDOM. The authors gratefully acknowledge the numerical support provided by Dr. J. S. H Leeuwaarden (EURANDOM) and his remark leading to the last five lines of Section IV.

## REFERENCES

[1] D. Anick, D. Mitra and M. M. Sondhi, *Stochastic theory of a data handling system with multiple sources*, Bell Syst. Tech. J., Vol.61:1871-1894, 1982.  
[2] S. Asmussen, *Applied Probability and Queues*, Springer, 2nd edition, 2003.  
[3] S. Asmussen, C. Klüppelberg, and K. Sigman, *Sampling at subexponential times with Queueing Applications*, Stoch. Proc. and Appl., Vol. 79: 265-286, 1999.

[4] F. Baccelli, S. Schlegel and V. Schmidt, *Asymptotics of stochastic networks with subexponential service times*, Queueing Systems, Vol. 33: 205-232, 1999.  
[5] N. H. Bingham, C. M. Goldie and J. L. Teugels, *Regular Variation*, Cambridge Univ. Press, 1987.  
[6] O. J. Boxma, *Fluid queues and regular variation*, Performance Evaluation, Vol. 27 & 28: 699-712, 1996.  
[7] H. Bruneel and B. G. Kim, *Discrete-Time Models for Communication Systems including ATM*, Kluwer Academic Publishing, Boston, 1993.  
[8] J. W. Cohen, *The Single Server Queue*, North-Holland Publishing Company, Amsterdam, 2nd edition, 1982.  
[9] H. Daduna, *Queueing Networks with Discrete-Time Scale*, LNCS, Springer, 2001.  
[10] A. Eryilmaz, R. Srikant and J. R. Perkins, *Stable Scheduling policies for fading wireless channels*, IEEE/ACM Trans. Networking, Vol. 13:411-424, 2005.  
[11] A. M. Greiner, M. Jobmann and C. Klüppelberg, *Telecommunication traffic, queueing models and subexponential distributions*, Queueing Systems, Vol. 33:125-152, 1999.  
[12] A. Gut, *Stopped Random Walk*, Springer, N. Y., 1991.  
[13] T. Huang and K. Sigman, *Steady-state asymptotics for tandem, split match and other feedforward queues with heavy tailed services*, Queueing Systems, Vol. 33:233-259, 1999.  
[14] J. Kiefer and J. Wolfowitz, *On the characterization of the general queueing process with applications to random walks*, Annals Math. Stat, Vol. 27:147-161, 1956.  
[15] M. Kodialam and T. Nandogopal, *Characterizing the capacity region in multi-radio, multi-channel wireless mesh networks*, Proc. ACM Conf. Mobicom, 2005.  
[16] L. Kosten, *Stochastic theory of data handling systems with groups of multiple sources*, pages 321-331, In Performance of Computer Communication Systems, W. Bux, H. Rudin (eds.), North Holland, Amsterdam, 1984.  
[17] W. Krämer and M. Langenbach-Beltz, *Approximate formulae for the delay in the queueing system GI/G/1*, Eighth International Telemetric Congress, Melbourne, pages 235-1-8, 1776.  
[18] J. S. H. Van Leeuwaarden, *Queueing Models for Cable Access Networks*, Ph.D Thesis, Technische Universiteit Eindhoven, The Netherlands, 2005.  
[19] R. M. Loynes, *The Stability of a queue with nonindependent interarrival and service time distributions*, Proc. Cambridge Phil. Society, Vol. 58:479-520, 1962.  
[20] G. Mergen and L. Tong, *Stability and Capacity of regular wireless networks*, IEEE Trans. Inf. Theory, Vol. 51:1938-1953, 2005.  
[21] M. J. Neely, E. Modiano and C-P Li, *Fairness and optimal stochastic control for heterogeneous networks*, in. Proc. Conf. INFOCOM, 2005.  
[22] M. J. Neely, E. Modiano and C. E. Rihors, *Dynamic power allocation and routing for time varying wireless networks*, IEEE INFOCOM, Vol. 1: 745-755, 2003.  
[23] F. Ohrtman, *WiMAX Handbook: Building 802.16 Wireless Networks*, McGraw-Hill, Professional, 2005.  
[24] B. Radunović and J-Y LeBoudec, *Rate performance objectives of multihop wireless networks*, IEEE Trans. Mobile Computing, Vol. 3:334-349, 2004.  
[25] J. W. Roberts, *Performance Evaluation and Design of Multiserver networks*, COST Project 224, Final report, 1991.  
[26] R. Rom and M. Sidi, *Message delay distribution in generalized time division multiple access*, Prob. Eng. Inform. Sci. Vol 4: 187-202, 1990.  
[27] I. Rubin and Z. Zhang, *Message delay analysis for TDMA schemes using contiguous slot assignments*, IEEE Trans. Communications, Vol 40:730-737, 1992.  
[28] V. Sharma, *Some limit theorems for regenerative queues*, Queueing Systems, Vol. 30:341-363, 1998.  
[29] V. Sharma and N. D. Gangadhar, *Asymptotics for transient and stationary probabilities for finite and infinite buffer discrete-time queues*, Queueing Systems, Vol. 26:1-22, 1997.  
[30] K. Sigman, *A primer on heavy tailed distributions*, Queueing Systems, Vol. 33: 261-275, 1999.