A tandem queue with a gate mechanism

Zbigniew Palmowski[1,2,3] Sabine Schlegel[1] Onno Boxma[1,4]

January 28, 2002

Abstract

Inspired by a problem regarding cable access networks, we consider a two station tandem queue with Poisson arrivals. At station 1 we operate a gate mechanism, leading to batch arrivals at station 2. Upon arrival at station 1, customers join a queue in front of a gate. Whenever all customers present at the service area of station 1 have received service, the gate before as well as a gate behind the service facility open. Customers leave the service area and enter station 2 (as a batch), while all customers waiting at the gate in front of station 1 are admitted into the service area. For station 1 we analyse the batch size and the time between two successive gate openings, as well as waiting and sojourn times of individual customers for different service disciplines. For station 2, we investigate waiting times of batch customers, where we allow that service times may depend on the size of the batch and also on the interarrival time. In the analysis we use Wiener-Hopf factorization techniques for Markov modulated random walks.

Keywords: tandem queue, gate mechanism, batch customers, access networks, collision resolution

2000 Mathematics Subject Classification: 60K25, 60K20, 60K15, 60K30, 68M20, 90B22, 47A68.

1 Introduction

Emerging access networks like residential cable networks with burst speeds up to 40 Mb/s - more than 2000 times faster than an ordinary dial-up modem - will most likely provide the next-generation data communication services, including internet access to homes and small business. These networks are currently being standardized (e.g. DOCSIS, IEEE, DAVIC/DVB) and therefore they are the focus of extensive research activity. Characteristic for access networks is a two-stage sequential procedure for data transfer from a station at the customers' premises to a central node. First, a contention stage is carried out (with other stations), in which stations that have data to transmit request a number of data slots. This is done by growing a Capetanakis type contention tree [8]. Requests that are

[1]EURANDOM, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
[2]Mathematical Institute, University of Wrocław, pl. Grunwaldzki 2/4, 50-384 Wrocław, Poland
[3]Work supported by KBN under grant 5 P03A 02120
[4]Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

successfully received by the central node enter the second stage: they queue in a request queue, until they are scheduled by a centralized scheduler. When scheduled, the station can transmit the data collision-free.

Performance analysis of access networks has mainly been investigated via simulation, see Limb and Sala [15]. Self regulation in access networks has been studied e.g. in Kelly [14], Örmeci and Resing [21]; see also Cohen [10]. Contention trees are extensively analysed; see, e.g., Flajolet and Mathys [12], Capetanakis [8] and Tsybakov and Mikhailov [28]. The literature on queueing systems useful for modelling access networks is still small. This is due to a number of complications which arise from dependencies between the contention tree resolution and the data transmission stage.

In this paper we propose and analyse the so-called "gated model" for access networks. This model consists of two single server queues Q_1 and Q_2 in tandem. At Q_1 we operate a gate mechanism which leads to batch arrivals at Q_2. Q_1 represents the collision resolution procedure in the Capetanakis-Tsybakov-Mikhailov contention tree algorithm, i.e. the procedure for determining the order of sending requests for data transmission, whereas Q_2 then represents data transmission.

In Section 2 we describe the model in detail. In Section 3 we find a number of quantities related to Q_1: the number of customers present at the gate opening, the length of the gate period, sojourn and waiting times. In Section 4 we analyse Q_2. In particular, we find the Laplace-Stieltjes transform of the steady-state waiting time distribution at Q_2.

2 Model description and notation

We consider a tandem of two single server queues Q_1 and Q_2. Customers arrive at Q_1 according to a Poisson process with intensity $\lambda > 0$. Upon arrival, they join a queue. Entrance and departure from the service area of Q_1 is regulated by the following gate mechanism. After being served at Q_1, customers enter an infinite waiting room behind the service facility. Assume that there is a gate both behind that waiting room and in front of the service facility. Whenever all customers present in the service area of Q_1 have received service, both gates open simultaneously. Customers from the waiting room leave (as a batch) and enter Q_2. All customers waiting at the gate in front of Q_1 are admitted into the service area and the gates close immediately. If there are no waiting customers, then the gates remain open and are closed immediately after the first arrival. Service at Q_1 is as follows: when there are G_n customers present at the moment that the gate opens for the n^{th} time, then these G_n customers receive service, which takes a stochastic amount of time $R(G_n)$. Customers are not necessarily served in order of arrival, and may not even be served one by one. However, in this paper we assume that each customer receives individual service and that

$$\begin{cases} R(G_n) &= B_1 + \ldots + B_{G_n}, \quad \text{if } G_n > 0, \\ R(0) &= B_1, \end{cases} \tag{2.1}$$

where the B_i are i.i.d. random variables (r.v.) with distribution $B(\cdot)$, with mean β, second moment $\beta^{(2)}$ and Laplace-Stieltjes Transform (LST) $\beta(s)$. This approach to model $R(G_n)$ is motivated by the observation that the mean and the variance of the time needed to resolve a contention tree are asymptotically linear in G_n as $n \to \infty$; see Denteneer and Pronk [11] and Capetanakis [8]; see also Flajolet [12] and Janssen [13]. (Batch) Service times at Q_2 may depend on the batch size as well as on the interarrival times and will be discussed in detail in Section 4.

3 Analysis of Q_1

In this section we first analyse the number of customers present at gate openings, i.e. the batch size of arrivals to Q_2, and the gate periods, i.e. the interarrival times at Q_2. This then enables us to investigate waiting times of individual customers.

3.1 The number at the gate

Let $A(x)$ denote the number of Poisson(λ) arrivals during a period of length x. Then, from (2.1) we have

$$\begin{cases} G_{n+1} &= A(B_1 + \ldots + B_{G_n}), \quad \text{if } G_n > 0, \\ G_{n+1} &= A(B_1), \quad \text{if } G_n = 0. \end{cases}$$

Hence, with $\mathbb{E}[z^{A(B_1)}] = \beta(\lambda(1-z))$,

$$\begin{aligned} \mathbb{E}[z^{G_{n+1}}] &= \mathbb{E}[z^{A(B_1+\ldots+B_{G_n})}] - P(G_n = 0) + P(G_n = 0)\mathbb{E}[z^{A(B_1)}] \\ &= \mathbb{E}[\beta(\lambda(1-z))^{G_n}] - P(G_n = 0) + P(G_n = 0)\beta(\lambda(1-z)). \end{aligned} \tag{3.1}$$

As far as workload is concerned, the gate in front of Q_1 has no influence since it does not matter whether customers wait directly in front of the server or whether there is a gate in between. Therefore, the total amount of work at Q_1 behaves exactly like the total amount of work in the corresponding $M/G/1$ queue without gate, and the steady state distribution of the workload exists if and only if $\rho := \lambda\beta < 1$. The latter statement also holds for the number of customers at gate openings. In the sequel we assume the traffic load ρ to be less than one. Denote by G the r.v. with distribution the steady state distribution of G_n, and let $G(z)$ denote the Generating Function (GF) of G. It follows from (3.1) that, for $|z| \leq 1$,

$$G(z) = [G(\beta(\lambda(1-z))) - G(0)] + G(0)\beta(\lambda(1-z)). \tag{3.2}$$

With the notation $f(z) := \beta(\lambda(1-z))$, this becomes, for $|z| \leq 1$,

$$G(z) = G(f(z)) - G(0)(1 - f(z)), \tag{3.3}$$

and, after one iteration,

$$G(z) = G(f(f(z))) - G(0)(1 - f(f(z))) - G(0)(1 - f(z)).$$

Introduce, for $|z| \leq 1$,

$$f_j(z) := \beta(\lambda(1 - f_{j-1}(z))) = f(f_{j-1}(z)), \quad j = 1, 2, \ldots,$$

with $f_0(z) \equiv z$. Then, after k iterations, (3.3) results in

$$G(z) = G(f_k(z)) - G(0) \sum_{j=1}^{k} [1 - f_j(z)].$$

Following an argument in Boxma and Cohen [7], it is easily seen that $f_k(z) \to 1$ for $k \to \infty$ if $\rho < 1$, and the convergence is geometrically fast so that $\sum_{j=1}^{\infty}[1 - f_j(z)]$ converges. Using $G(1) = 1$ we obtain

$$G(z) = 1 - G(0) \sum_{j=1}^{\infty} [1 - f_j(z)], \tag{3.4}$$

and finally

$$G(0) = \frac{1}{1 + \sum_{j=1}^{\infty}[1 - f_j(0)]}.$$

From (3.4) or directly from (3.2) we find via differentiation

$$\mathbb{E}G = G'(1) = \frac{\rho}{1-\rho} G(0), \quad \mathbb{E}G^2 = \frac{G(0)}{1-\rho}\Big(\rho + \frac{\lambda^2 \beta^{(2)}}{1-\rho^2}\Big).$$

3.2 The gate period

The steady state distribution $D(t) = P(D < t)$ of the time D between two successive gate openings, *which is the steady state interarrival time distribution at* Q_2, directly follows from (2.1). Clearly, with $*$ denoting a convolution,

$$D(t) = \sum_{k=1}^{\infty} P(G = k) B^{*k}(t) + P(G = 0)(1 - \mathrm{e}^{-\lambda t}) * B(t). \tag{3.5}$$

Hence the LST is given by

$$\begin{aligned} \mathbb{E}[\mathrm{e}^{-sD}] &= G(\beta(s)) - G(0) + G(0) \frac{\lambda}{\lambda + s} \beta(s) \\ &= 1 - G(0)[\sum_{j=1}^{\infty}[1 - f_j(\beta(s))] + 1 - \frac{\lambda}{\lambda + s}\beta(s)], \quad \mathrm{Re}\, s \geq 0. \end{aligned} \tag{3.6}$$

In particular, from (3.6) or directly relating $\mathbb{E}D$ and $\mathbb{E}G$, we get

$$\mathbb{E}D = \frac{\beta G(0)}{\rho(1-\rho)}, \quad \mathbb{E}D^2 = G(0)\Big(\frac{\beta^{(2)}}{(1-\rho)(1-\rho^2)} + \frac{2(1+\rho)}{\lambda^2}\Big). \tag{3.7}$$

3.3 Waiting times

We now investigate the stationary waiting time W_1 of an individual customer at Q_1. We consider two different service disciplines. For First-Come-First-Served (FCFS) the gate mechanism does not matter and W_1 behaves like in an ordinary $M/G/1$ queue. Thus, the LST of W_1 is given by [5]

$$\mathbb{E}\mathrm{e}^{-sW_1} = \frac{(1-\rho)s}{s-\lambda+\lambda\beta(s)}\,,$$

and in particular,

$$\mathbb{E}W_1 = \frac{\rho}{1-\rho}\,\frac{\beta^{(2)}}{2\beta} = \frac{\lambda\beta^{(2)}}{2(1-\rho)}\,.$$

Alternatively, the waiting time can be considered as follows. Given that a customer finds a non-empty system, his waiting time is the sum of the residual time until the following gate opening and the service times of those customers who are served before him and who belong to the same batch of customers. We use this approach to compute the mean waiting time $\mathbb{E}W_1^{\mathrm{ROS}}$ in the case of random order of service (ROS) within a batch. By this we mean that on entering the service area a batch is ordered randomly, where each ordering has equal probability, and is then served according to this order. Thus, the probability of being served i^{th} in a batch of size k is given by $1/k$. Note that there are two types of intervals between two successive gate openings (we call these intervals *gating periods*). If there are no customers waiting when the gates open, then the following gating period, denoted by D^0, consists of a residual interarrival time of the Poisson arrival process and exactly one service time B, having distribution

$$D^0(t) = P(D^0 < t) = (1-\mathrm{e}^{-\lambda t}) * B(t)\,.$$

Otherwise the gating period is given by the sum of G i.i.d. service times, i.e. $D^* = B_1 + \ldots + B_G$, and it has distribution

$$D^*(t) = P(D^* < t) = \sum_{k=1}^{\infty} P(G=k)B^{*k}(t)/(1-G(0))\,. \tag{3.8}$$

Since a gating period of type D^0 occurs if and only if Q_1 is empty, the probability $G^0(0)$ for a customer to arrive during such a period is given by

$$G^0(0) = \frac{\lambda^{-1}+\beta}{\lambda^{-1}+\mathbb{E}B_{\mathrm{busy}}} = 1-\rho^2\,,$$

where B_{busy} denotes the busy period in the corresponding $M/G/1$ queue. Thus, with the notation

$$\begin{aligned} p_k &= P\Big(A(B^{\mathrm{past}}) + A(B^{\mathrm{res}}) + 1 = k\Big)\,, \\ p_k^* &= P\Big(A(D^{*\mathrm{past}}) + A(D^{*\mathrm{res}}) + 1 = k\Big)\,, \end{aligned}$$

where X^{res} and X^{past} denote the forward and backward recurrence time of a r.v. X, we get

$$
\begin{aligned}
\mathbb{E}W_1^{\mathrm{ROS}} &= G^0(0)\,\mathbb{E}[W_1^{\mathrm{ROS}}|D^0] + (1-G^0(0))\,\mathbb{E}[W_1^{\mathrm{ROS}}|D^*] \\
&= G^0(0)\Big[\frac{1-\rho}{G^0(0)}\cdot 0 + \Big(1-\frac{1-\rho}{G^0(0)}\Big)\Big(\frac{\beta^{(2)}}{2\beta} + \sum_{k=1}^{\infty}\sum_{i=1}^{k}\sum_{l=1}^{i-1}\mathbb{E}B_l\,\frac{1}{k}\,p_k\Big)\Big] \\
&\quad +(1-G^0(0))\Big[\frac{\mathbb{E}D^{*2}}{2\mathbb{E}D^*} + \sum_{k=1}^{\infty}\sum_{i=1}^{k}\sum_{l=1}^{i-1}\mathbb{E}B_l\,\frac{1}{k}\,p_k^*\Big], \qquad (3.9)
\end{aligned}
$$

and it remains to calculate p_k and p_k^*. We have

$$
\begin{aligned}
p_k &= \int_{t_p=0}^{\infty}\int_{t_r=0}^{\infty} P(A(t_p+t_r)=k-1)\,\mathrm{d}P(B^{\mathrm{past}}<t_p, B^{\mathrm{res}}<t_r) \\
&= \int_{t_p=0}^{\infty}\int_{t_r=0}^{\infty} \mathrm{e}^{-\lambda(t_p+t_r)}\,\frac{(\lambda(t_p+t_r))^{k-1}}{(k-1)!}\,\mathrm{d}P(B^{\mathrm{past}}<t_p, B^{\mathrm{res}}<t_r)\,.
\end{aligned}
$$

Let $b(\cdot)$ denote the density of $B(\cdot)$. Then, see e.g. Takagi [27], p.17,

$$
\begin{aligned}
p_k &= \frac{1}{\beta}\int_{t_p=0}^{\infty}\int_{t_r=0}^{\infty} \mathrm{e}^{-\lambda(t_p+t_r)}\,\frac{(\lambda(t_p+t_r))^{k-1}}{(k-1)!}\,b(t_p+t_r)\,\mathrm{d}t_r\,\mathrm{d}t_p \\
&= \frac{1}{\beta}\,\frac{1}{(k-1)!}\,\frac{1}{\lambda}\int_{x=0}^{\infty}\mathrm{e}^{-\lambda x}(\lambda x)^k b(x)\,\mathrm{d}x\,. \qquad (3.10)
\end{aligned}
$$

Analogously, with $d^*(\cdot)$ denoting the density of the distribution of D^*,

$$
p_k^* = \frac{1}{\mathrm{E}D^*}\,\frac{1}{(k-1)!}\,\frac{1}{\lambda}\int_{x=0}^{\infty}\mathrm{e}^{-\lambda x}(\lambda x)^k\,d^*(x)\,\mathrm{d}x\,. \qquad (3.11)
$$

Substituting the expressions in (3.10) and (3.11) into (3.9) yields

$$
\begin{aligned}
\mathbb{E}W_1^{\mathrm{ROS}} &= (\rho-\rho^2)\Big(\frac{\beta^{(2)}}{2\beta} + \sum_{k=1}^{\infty}\sum_{i=1}^{k}(i-1)\beta\,\frac{1}{k\beta(k-1)!}\,\frac{1}{\lambda}\int_0^{\infty}\mathrm{e}^{-\lambda x}(\lambda x)^k\,b(x)\,\mathrm{d}x\Big) \\
&\quad +\rho^2\Big(\frac{\mathbb{E}D^{*2}}{2\mathbb{E}D^*} + \sum_{k=1}^{\infty}\sum_{i=1}^{k}(i-1)\beta\,\frac{1}{k\mathbb{E}D^*(k-1)!}\,\frac{1}{\lambda}\int_0^{\infty}\mathrm{e}^{-\lambda x}(\lambda x)^k\,d^*(x)\,\mathrm{d}x\Big) \\
&= (\rho-\rho^2)\Big(\frac{\beta^{(2)}}{2\beta} + \frac{1}{2\lambda}\int_0^{\infty}\mathrm{e}^{-\lambda x}(\lambda x)^2\sum_{k=2}^{\infty}\frac{(\lambda x)^{k-2}}{(k-2)!}\,b(x)\,\mathrm{d}x\Big) \\
&\quad +\rho^2\Big(\frac{\mathbb{E}D^{*2}}{2\mathbb{E}D^*} + \frac{\beta}{2\lambda\mathbb{E}D^*}\int_0^{\infty}\mathrm{e}^{-\lambda x}(\lambda x)^2\sum_{k=2}^{\infty}\frac{(\lambda x)^{k-2}}{(k-2)!}\,d^*(x)\,\mathrm{d}x\Big) \\
&= (\rho-\rho^2)\Big(\frac{\beta^{(2)}}{2\beta} + \frac{\rho\beta^{(2)}}{2\beta}\Big) + \rho^2\Big(\frac{\mathbb{E}D^{*2}}{2\mathbb{E}D^*} + \frac{\rho\mathbb{E}D^{*2}}{2\mathbb{E}D^*}\Big) = \frac{\rho}{1-\rho}\,\frac{\beta^{(2)}}{2\beta}\,,
\end{aligned}
$$

where in the last step we have used

$$
\frac{\mathbb{E}D^{*2}}{2\mathbb{E}D^*} = \frac{\beta^{(2)}(1+\rho-\rho^2)}{2\beta(1-\rho^2)}\,,
$$

which can be seen as follows. From (3.8) we get

$$
\mathbb{E}D^* = \sum_{k=1}^{\infty} P(G=k)k\beta/(1-G(0)) = \beta\mathbb{E}G/(1-G(0))
$$

and

$$\begin{aligned}
\mathbb{E}D^{*2} &= \sum_{k=1}^{\infty} P(G=k)\mathbb{E}[(B_1+\ldots+B_k)^2]/(1-G(0)) \\
&= \sum_{k=1}^{\infty} P(G=k)\Big(k(\beta^{(2)}-\beta^2)+(k\beta)^2\Big)/(1-G(0)) \\
&= \Big((\beta^{(2)}-\beta^2)\mathbb{E}G+\beta^2\mathbb{E}G^2\Big)/(1-G(0))\ .
\end{aligned}$$

Note that the mean waiting time is the same for FCFS and ROS. In the same way we also get the LST of the waiting time:

$$\begin{aligned}
\mathbb{E}\mathrm{e}^{-\omega W_1^{\mathrm{ROS}}} &= G^0(0)\mathbb{E}\Big[\mathrm{e}^{-\omega W_1^{\mathrm{ROS}}}|D^0\Big]+(1-G^0(0))\mathbb{E}\Big[\mathrm{e}^{-\omega W_1^{\mathrm{ROS}}}|D^*\Big] \\
&= 1-\rho+(\rho-\rho^2)\int_{x=0}^{\infty}\frac{1-\mathrm{e}^{-\lambda(1-\beta(\omega))x}}{\lambda(1-\beta(\omega))x}\,\frac{1-\mathrm{e}^{-\omega x}}{x}\,\frac{b(x)}{\beta}\,\mathrm{d}x \\
&\quad +\rho^2\int_{x=0}^{\infty}\frac{1-\mathrm{e}^{-\lambda(1-\beta(\omega))x}}{\lambda(1-\beta(\omega))x}\,\frac{1-\mathrm{e}^{-\omega x}}{x}\,\frac{d^*(x)}{\mathbb{E}D^*}\,\mathrm{d}x\ .
\end{aligned}$$

3.4 Sojourn times

We now analyse the (stationary) sojourn time $S^{(1)}$ of an individual customer in Q_1, i.e., the time from arrival to the system to departure from Q_1 or, which is equivalent, entrance to Q_2. Other than the waiting time, the sojourn time does not behave like in the corresponding $M/G/1$ queue without gates. Namely, due to the gate mechanism, a customer's sojourn time consists of the residual gating period (unless the customer finds an empty system upon arrival), his own service time, and the service times of the customers arriving during both the backward and forward gating period, i.e.,

$$S^{(1)} = \begin{cases} B & \text{w.p. } 1-\rho, \\ B^{\mathrm{res}}+B+\sum_{k=1}^{A(B^{\mathrm{past}})}B_{2k}+\sum_{k=1}^{A(B^{\mathrm{res}})}B_{2k+1} & \text{w.p. } \rho-\rho^2, \\ D^{*\mathrm{res}}+B+\sum_{k=1}^{A(D^{*\mathrm{past}})}B_{2k}+\sum_{k=1}^{A(D^{*\mathrm{res}})}B_{2k+1} & \text{w.p. } \rho^2. \end{cases} \tag{3.12}$$

From this representation we obtain the LST

$$\begin{aligned}
\mathbb{E}\mathrm{e}^{-\omega S^{(1)}} &= (1-\rho)\mathbb{E}\mathrm{e}^{-\omega B}+(\rho-\rho^2)\mathbb{E}\Big[\mathrm{e}^{-\omega(B^{\mathrm{res}}+B+\sum_{k=1}^{A(B^{\mathrm{past}}+B^{\mathrm{res}})}B_k)}\Big] \\
&\quad +\rho^2\mathbb{E}\Big[\mathrm{e}^{-\omega(D^{*\mathrm{res}}+B+\sum_{k=1}^{A(D^{*\mathrm{past}}+D^{*\mathrm{res}})}B_k)}\Big] \\
&= (1-\rho)\mathbb{E}\Big[\mathrm{e}^{-\omega B}\Big]+(\rho-\rho^2)\mathbb{E}\Big[\mathrm{e}^{-\omega B}\Big]\mathbb{E}\Big[\mathrm{e}^{-\omega(B^{\mathrm{res}}+\sum_{k=1}^{A(B^{\mathrm{past}}+B^{\mathrm{res}})}B_k)}\Big] \\
&\quad +\rho^2\mathbb{E}\Big[\mathrm{e}^{-\omega B}\Big]\mathbb{E}\Big[\mathrm{e}^{-\omega(D^{*\mathrm{res}}+\sum_{k=1}^{A(D^{*\mathrm{past}}+D^{*\mathrm{res}})}B_k)}\Big] \\
&= (1-\rho)\beta(\omega)+(\rho-\rho^2)\beta(\omega)\hat{\beta}(\omega)+\rho^2\beta(\omega)\hat{D}^*(\omega)\ ,
\end{aligned}$$

where

$$\hat{\beta}(\omega)=\mathbb{E}\Big[\mathrm{e}^{-\omega(B^{\mathrm{res}}+B_1+\ldots+B_{A(B^{\mathrm{past}}+B^{\mathrm{res}})})}\Big]$$

$$
\begin{aligned}
&= \int_{t_p=0}^{\infty}\int_{t_r=0}^{\infty} \mathrm{e}^{-\omega t_r} \int_{t=0}^{\infty} \mathrm{e}^{-\omega t} \mathrm{d}P(B_1 + \ldots + B_{A(t_p+t_r)} < t) \mathrm{d}P(B^{\text{past}} < t_p, B^{\text{res}} < t_r) \\
&= \int_{t_p=0}^{\infty}\int_{t_r=0}^{\infty} \mathrm{e}^{-\omega t_r} \mathrm{e}^{-\lambda(t_p+t_r)(1-\beta(\omega))} \mathrm{d}P(B^{\text{past}} < t_p, B^{\text{res}} < t_r) \\
&= \frac{\beta\Big(\lambda(1-\beta(\omega))\Big) - \beta\Big(\omega + \lambda(1-\beta(\omega))\Big)}{\omega\beta} ,
\end{aligned}
$$

and, analogously,

$$
\begin{aligned}
\hat{D}^*(\omega) &= \mathbb{E}\Big[\mathrm{e}^{-\omega(D^{*\text{res}} + B_1 + \ldots + B_{A(D^{*\text{past}} + D^{*\text{res}})})}\Big] \\
&= \frac{\tilde{D}^*\Big(\lambda(1-\beta(\omega))\Big) - \tilde{D}^*\Big(\omega + \lambda(1-\beta(\omega))\Big)}{\omega \mathbb{E} D^*}
\end{aligned}
$$

with $\tilde{D}^*$ denoting the LST of D^* given by $\tilde{D}^*(s) = (G(\beta(s)) - 1)/(1 - G(0))$, see (3.8). From (3.2) we have

$$
\begin{aligned}
& G\Big(\beta(\omega + \lambda(1-\beta(\omega)))\Big) - G\Big(\beta(\lambda(1-\beta(\omega)))\Big) \\
&\quad = G\Big(\beta(\omega) - \frac{\omega}{\lambda}\Big) - G\Big(\beta(\omega)\Big) \\
&\qquad\qquad + G(0)\Big[\beta\Big(\lambda(1-\beta(\omega))\Big) - \beta\Big(\omega + \lambda(1-\beta(\omega))\Big)\Big] ,
\end{aligned}
$$

and thus, finally,

$$
\mathbb{E}\mathrm{e}^{-\omega S^{(1)}} = \beta(\omega)(1-\rho)\Big[1 + \lambda \frac{G(\beta(\omega)) - G(\beta(\omega) - \omega/\lambda)}{\omega G(0)}\Big] .
$$

Via differentiation, or directly from the representation (3.12), we obtain

$$
\mathbb{E} S^{(1)} = \frac{\lambda \beta^{(2)}(1+2\rho)}{2(1-\rho^2)} + \beta .
$$

4 Analysis of Q_2

In this section we analyse the waiting time of a batch customer at Q_2. The stationary waiting time W_2 of a batch customer is equal in distribution to the supremum of a Markov modulated random walk (MMRW), where the governing Markov chain (MC) has countable state space. We show that this MMRW can be approximated by a MMRW governed by a MC with finite state space in the sense that the waiting time in the finite state space model converges in distribution to the waiting time in the original model. Using Wiener-Hopf type arguments we derive the LST of the stationary waiting time in the finite state space model. Before we do so, we give a short description of the model and discuss the modelling of batch service times.

The arrival process to Q_2 is the departure process from Q_1, i.e. the interarrival times are the gating periods D with distribution

$$
D(t) = \sum_{k=1}^{\infty} P(G=k) B^{*k}(t) + P(G=0)(1 - \mathrm{e}^{-\lambda t}) * B(t), \tag{4.1}
$$

and mean

$$\mathbb{E}D = \frac{\beta G(0)}{\rho(1-\rho)} ,$$

see (3.5) and (3.7). We assume that the system is in steady state and that the n^{th} batch customer requires a service

$$B_n^{(2)} = C_0 \mathbb{I}(G_n = 0) + \left(\sum_{k=1}^{G_n} C_k + aD_n \right) \mathbb{I}(G_n > 0) , \tag{4.2}$$

with some constant $a \geq 0$, where $\{C_n\}$ are independent exponentially distributed r.v.'s with mean c, D_n denotes the n^{th} gating period, and G_n is the size of the n^{th} batch. By $B^{(2)}$ we denote a generic r.v. $B_n^{(2)}$. Note that successive service times and interarrival times are both governed by G_n, so both are *not* i.i.d. This approach to model batch service times at Q_2 is motivated by the two-stage transmission mechanism in access networks as described in Section 1. Before actually sending data, a station has to request the number of slots required to send these data, which takes place in contention with other sources. During the resolution of possible collisions stations may increase the number of requested slots (*updating*). In (4.2), the C_n's represent the original amount of data, and the constant a reflects the updating of requests during collision resolution. We are interested in the steady state waiting time distribution. This will exist if $\mathbb{E}B^{(2)} < \mathbb{E}D$, i.e. if

$$\begin{aligned} \mathbb{E}B^{(2)} &= cP(G=0) + \sum_{n=1}^{\infty} \mathbb{E}\Big[\sum_{k=1}^{n} C_k + a(B_1 + \ldots + B_n)\Big] P(G=n) \\ &= \frac{G(0)}{1-\rho}(c + a\beta\rho) < \frac{\beta G(0)}{\rho(1-\rho)} , \end{aligned}$$

or, equivalently,

$$c + a\beta\rho < \frac{1}{\lambda} , \tag{4.3}$$

hence for

$$\lambda < \lambda_0 := \frac{-c + \sqrt{c^2 + 4a\beta^2}}{2a\beta^2} .$$

4.1 Waiting times at Q_2

Let $W_{2,n}$ denote the actual waiting time of the n^{th} batch customer at Q_2. Let

$$S_0 = 0, \qquad S_n = \sum_{k=1}^{n} X_k \ \ (n \geq 1) ,$$

where

$$X_{k+1} = B_k^{(2)} - D_k .$$

Then $\{(S_n, G_n), n \in \mathbb{N}\}$ is a MMRW. The governing MC $\{G_n\}$ has state space $\{0, 1, 2, \ldots\}$. By $P = (p_{ij})$ we denote its transition matrix. As in the classical case, we have

$$W_{2,0} = 0 , \quad W_{2,n} = S_n - \min_{0 \leq k \leq n} S_k .$$

We now specify the distribution $B(\cdot)$ to be an Erlang-κ distribution with mean β,

$$B \stackrel{d}{=} \mathrm{Erl}(\kappa, \kappa/\beta) .$$

This choice is motivated by results from Denteneer and Pronk [11].

Let $H_i(\cdot)$ $(i \geq 1)$ denote the distribution function of $\sum_{k=1}^{i} C_k$, that is

$$H_i \stackrel{d}{=} \mathrm{Erl}(i, 1/c) \ .$$

Then, for $i \in \{1, 2, \ldots\}$ and $j \in \{0, 1, 2 \ldots\}$ it holds that

$$\begin{aligned} & P(\sum_{l=1}^{i} B_l \leq x, \sum_{l=1}^{i} C_l \leq y, G_k = j | G_{k-1} = i) \\ = & \int_0^x \frac{(\kappa/\beta)^{\kappa i}}{(\kappa i - 1)!} w^{\kappa i - 1} e^{-w\kappa/\beta} \frac{(\lambda w)^j}{j!} \mathrm{e}^{-\lambda w} \, \mathrm{d}w H_i(y) \\ = & \ p_{ij} F_{ij}(x) H_i(y) \ , \end{aligned}$$

where

$$p_{ij} = \binom{\kappa i + j - 1}{j} \left(\frac{\kappa/\beta}{\kappa/\beta + \lambda} \right)^{\kappa i} \left(\frac{\lambda}{\kappa/\beta + \lambda} \right)^{j}$$

is the transition probability $p_{ij} = P(G_k = j | G_{k-1} = i)$ of the underlying MC $\{G_n\}$, and

$$F_{ij} \stackrel{d}{=} \mathrm{Erl}(\kappa i + j, \kappa/\beta + \lambda) .$$

Moreover,

$$P(D^0 \leq x, C_1 \leq y, G_k = j | G_{k-1} = 0) = p_{0j} F_{0j}(x) H_1(y) \ ,$$

where $p_{0j} = p_{1j}$ and $F_{0j}(x) = (1 - e^{-\lambda x}) * F_{1j}(x)$. For a distribution function $F(\cdot)$ and some constant $b < 1$ we define the distribution function $F^b(x)$ as

$$F^b(x) = 1 - F(-x/(1-b)) \ ,$$

i.e., if X has distribution F, then F^b is the distribution of $-(1-b)X$. With this notation, by (4.1) and (4.2) the transition kernel of the MMRW $\{(S_n, G_n)\}$ is given by

$$\begin{aligned} Q_{ij}(x) &= P(X_k \leq x, G_k = j | G_{k-1} = i) = p_{ij} H_i * F_{ij}^a(x), \qquad i \geq 1 , \\ Q_{0j}(x) &= P(X_k \leq x, G_k = j | G_{k-1} = 0) = p_{0j} H_1 * F_{0j}^0(x) \ . \end{aligned}$$

Denote by $\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots)$ the stationary distribution of the governing MC $\{G_n\}$ and by $\{(\hat{S}_n, \hat{G}_n), n \in \mathbb{N}\}$ the time-reversed version of the given MMRW whose transition kernel is given by

$$\hat{Q}_{ij}(x) = \frac{\pi_j}{\pi_i} Q_{ji}(x) \ .$$

Then from Prabhu and Tang [22], Lemma 1 and Th. 4, the steady-state limit W_2 of $W_{2,n}$ exists and has distribution:

$$W_2 \stackrel{d}{=} \sup_n \hat{S}_n \ ,$$

where for $\hat{G}_0$ we take the steady-state version of $\{\hat{G}_n\}$. Consider now the MC $\{G_n^K\}_{n\in\mathbb{N}}$ with *finite* state space $\{0,1,2,\ldots,K\}$ and transition matrix $P^K=(p_{ij}^K)$, where

$$p_{ij}^K = p_{ij}/\sum_{j=0}^{K} p_{ij} \ .$$

Define the MMRW $\{(S_n^K, G_n^K), n\in\mathbb{N}\}$ by

$$S_0^K = 0, \qquad S_n^K = \sum_{k=1}^{n} X_k^K$$

with corresponding transition kernel

$$Q_{ij}^K(x) = P(X_k^K \le x, G_k^K = j | G_{k-1}^K = i) = p_{ij}^K H_i * F_{ij}^a(x), \quad i=1,\ldots,K; j=0,\ldots,K,$$

$$Q_{0j}^K(x) = P(X_k^K \le x, G_k^K = j | G_{k-1}^K = 0) = p_{0j}^K H_1 * F_{0j}^0(x), \quad j=0,\ldots,K \ . \tag{4.4}$$

This means that we consider a Markov modulated queue in which conditioning on $\{G_{k-1}^K = i, G_k^K = j\}$, the interarrival and service times of the k^{th} customer have distribution $F_{ij}(x/(1-a))$ and $H_i(x)$, respectively, if $i>0$, and $F_{0j}(x)$ and $H_1(x)$ if $i=0$. Denote by $\boldsymbol{\pi}^K=(\pi_0^K,\ldots,\pi_K^K)$ the stationary distribution of the governing MC $\{G_n^K\}$, i.e. $\boldsymbol{\pi}^K$ solves

$$\boldsymbol{\pi}^K P^K = \boldsymbol{\pi}^K \ , \quad \sum_{i=0}^{K} \pi_i^K = 1 \ .$$

Further, consider the time-reversed version $\{(\hat{S}_n^K, \hat{G}_n^K), n\in\mathbb{N}\}$ of the given MMRW whose transition kernel is given by

$$\hat{Q}_{ij}^K(x) = \frac{\pi_j^K}{\pi_i^K} Q_{ji}^K(x) \ .$$

That is,

$$\hat{S}_0^K = 0\,, \qquad \hat{S}_n^K = \sum_{k=1}^{n} \hat{X}_k^K$$

and

$$P(\hat{X}_k^K \le x, \hat{G}_k^K = j | \hat{G}_{k-1}^K = i) = \frac{\pi_j^K}{\pi_i^K} P(X_k^K \le x, G_k^K = i | G_{k-1}^K = j) \ . \tag{4.5}$$

Then we have for the stationary waiting time in this queue

$$W_2^K \stackrel{d}{=} \sup_n \hat{S}_n^K \ ,$$

where for $\hat{G}_0^K$ we take the steady-state version of $\{\hat{G}_n^K\}$.

The following theorem allows us to approximate our model with countable state space by this model with finite state space.

Theorem 4.1 *The following convergence in distribution holds:*

$$W_2^K \xrightarrow{D} W_2 \qquad \text{as } K\to+\infty.$$

Proof. Let $\mathbb{E}_i^K$ and $\mathbb{E}_i$ denote the conditional expectations given that $\hat{G}_0^K = i$ and $\hat{G}_0 = i$, respectively. Then, the assertion of the theorem holds, if the MMRW $\{(\hat{S}_n, \hat{G}_n)\}$ is ergodic and

$$\mathbb{E}_i^K[\hat{X}_1^K; \hat{X}_1^K \geq 0 \wedge \hat{G}^K = j] \to \mathbb{E}_i[\hat{X}_1; \hat{X}_1 \geq 0 \wedge \hat{G} = j], \qquad \text{as } K \to \infty, \tag{4.6}$$

see Borovkov [6], Th. 22 and Th. 23, p. 53-54. Here $\mathbb{E}_i^K[\hat{X}_1^K; \hat{X}_1^K \geq 0 \wedge \hat{G}^K = j] = 0$ for $i > K$ or $j > K$. Condition (4.6) holds due to the choice of the kernel of the MMRW $\{(S_n^K, G_n^K), n \in \mathbb{N}\}$, whereas ergodicity of $\{(\hat{S}_n, \hat{G}_n)\}$ follows from regeneration arguments for the two-dimensional MC $\{(\hat{G}_{n-1}, \hat{G}_n), n \in \mathbb{N}\}$. Namely, following the lines of the proof of Asmussen [5], Th. 4.2, p. 235, we get

$$\frac{\hat{S}_n}{n} \to \boldsymbol{\pi}\boldsymbol{\nu} \qquad \text{a.s.} \; ,$$

where $\boldsymbol{\pi}$ is the stationary distribution of $\{G_n\}$ and $\boldsymbol{\nu}$ is the column vector with elements $\nu_i = \mathbb{E}_i[\hat{X}_1]$, $i = 0, \dots, K$. □

From now on we deal with the finite state space model for which we derive the LST $\mathbb{E}\exp\{-\omega W_2^K\}$. Following Prabhu and Tang [22], de Smit [24], de Smit and Regterschot [25], Arjas [2, 3], we first represent this quantity in terms of the LST of the ascending ladder heights associated with the time-reversed version of our MMRW. For this quantity an explicit expression can be obtained by solving a system of linear equations. Similar results for exponential interarrival or service times in $SM/SM/1$ can be found in e.g. Arjas [2], Neuts [20], Çinlar [9] and De Smit [24]. The analysis of waiting times in the $SM/SM/1$ model for Erlang interarrival or service times, as studied in this paper, still seems to be an open problem in general.

Let

$$\hat{T} = \min\{n > 0 : \hat{S}_n^K > 0\}, \qquad T = \min\{n > 0 : S_n^K \leq 0\} \, ,$$

denote the first ascending ladder epoch of $\{\hat{S}_n^K\}$ and the first descending ladder epoch of $\{S_n^K\}$, respectively. Finally, define the matrices $\chi^+(\omega)$ and $\chi^-(\omega)$ by

$$\chi^+(\omega) = \left\{\mathbb{E}_i(\exp(-\omega \hat{S}_{\hat{T}}^K), \hat{G}_{\hat{T}}^K = j)\right\}_{i,j=0}^{K} \quad , \qquad \operatorname{Re}\omega \geq 0 \, ,$$

and

$$\chi^-(\omega) = \left\{\mathbb{E}_i(\exp(-\omega S_T^K), G_T^K = j)\right\}_{i,j=0}^{K} \quad , \qquad \operatorname{Re}\omega \leq 0 \, .$$

Then, from Prabhu and Tang [22], Th. 4, the LST $\mathbb{E}_i(\exp(-\omega W_2^K), G^K = j)$ for $\operatorname{Re}\omega > 0$ is the j^{th} entry of the vector

$$\Pi[\mathbb{I} - \chi^+(\omega)]^{-1}[\mathbb{I} - \chi^+(0)]\boldsymbol{e} \; ,$$

and thus

$$\mathbb{E}_i(\exp(-\omega W_2^K)) = \sum_{j=0}^{K} \mathbb{E}_i(\exp(-\omega W_2^K), G^K = j) = \boldsymbol{e}'\Pi[\mathbb{I}-\chi^+(\omega)]^{-1}[\mathbb{I}-\chi^+(0)]\boldsymbol{e} \ , \quad (4.7)$$

where $\Pi = \mathrm{diag}(\boldsymbol{\pi}^K)$, $\mathbb{I} = \mathrm{diag}(\boldsymbol{e})$, and $\boldsymbol{e} = (1, \ldots, 1)'$. The prime denotes the transpose. In particular, $\mathbb{E}_i \exp(-\omega W_2^K)$ does not depend on the initial condition $G_0^K = i$. The remainder of this section is devoted to the derivation of an explicit expression for the terms on the right hand side of (4.7). To this end the following lemma is useful.

Lemma 4.1 (See Asmussen [5], Lem. 6.1, p. 216, for further generalizations.) *Let Z have an Erlang distribution* $\mathrm{Erl}(n, \theta)$ *and let L be a non-negative random variable which is independent of Z and has distribution F. Then, for $x \geq 0$, the density function of the random variable $Z - L$ is given by*

$$\sum_{m=0}^{n-1} \alpha_{n-m}(\theta) \frac{\theta^{m+1}}{m!} x^m \mathrm{e}^{-\theta x} \ , \quad (4.8)$$

where

$$\alpha_j(\theta) = \int_0^\infty \frac{(\theta y)^{j-1}}{(j-1)!} \mathrm{e}^{-\theta y} \mathrm{d}F(y) \ .$$

Proof. For $x \geq 0$, the density function of $Z - L$ is given by

$$\int_0^\infty \frac{\theta^n}{(n-1)!} (x+y)^{n-1} \mathrm{e}^{-\theta(x+y)} \mathrm{d}F(y) \ .$$

Expanding $(x+y)^{n-1}$ by the Binomial Theorem and rearranging terms yields (4.8). □

Let P_k denote the conditional probability given $G_0^K = k$. Then from (4.4) for $j > 0$,

$$\begin{aligned} & P_j(X_1^K \geq x+y, G_1^K = l | X_1^K \geq y) \\ = \ & \frac{P_j(X_1^K \geq x+y, G_1^K = l)}{P_j(X_1^K \geq y)} = \frac{p_{jl}^K - Q_{jl}^K(x+y)}{P_j(X_1^K \geq y)} = \frac{p_{jl}^K(1 - H_j * F_{jl}^a(x+y))}{P_j(X_1^K \geq y)} \\ = \ & \frac{p_{jl}^K}{P_j(X_1^K \geq y)} \left(\sum_{m=0}^{j-1} \alpha_{j-m}(1/c) \int_{x+y}^\infty \frac{(1/c)^{m+1}}{m!} w^m \mathrm{e}^{-w/c} \, \mathrm{d}w \right) \ , \end{aligned}$$

where the last step follows from Lemma 4.1. Hence, for $j > 0$,

$$\frac{\mathrm{d}}{\mathrm{d}x} P_j(X_1^K \geq x+y, G_1^K = l | X_1^K \geq y) = - \sum_{m=0}^{j-1} \sum_{i=0}^{m} a_{jl}^{(m,i)}(y) x^i \mathrm{e}^{-x/c} \quad (4.9)$$

for some functions $a_{jl}^{(m,i)}(y)$, $l, m, i \geq 0$, $j > 0$. Similarly,

$$\frac{\mathrm{d}}{\mathrm{d}x} P_0(X_1^K \geq x+y, G_1^K = l | X_1^K \geq y) = -a_l^{(0)}(y) x^i \mathrm{e}^{-x/c}$$

for some functions $a_l^{(0)}(y)$. Now

$$\chi_{kj}^+(\omega) = \int_0^\infty \sum_{n=1}^\infty \exp(-\omega x) \mathrm{d}g_{kj}^{(n)}(x) \, ,$$

where

$$\begin{aligned}
\mathrm{d}g_{kj}^{(n)}(x) &= \mathrm{d}P_k(\hat{S}_{\hat{T}}^K \le x, \hat{G}_{\hat{T}}^K = j, \hat{T} = n) \\
&= -\mathrm{d}P_k(\hat{S}_{\hat{T}}^K \ge x, \hat{G}_{\hat{T}}^K = j, \hat{T} = n) \\
&= -\sum_{l=0}^K \mathrm{d}P_k(\hat{S}_n^K \ge x, \hat{S}_1^K \le 0, \ldots, \hat{S}_{n-1}^K \le 0, \hat{S}_n^K > 0, \hat{G}_n^K = j, \hat{G}_{n-1}^K = l) \\
&= -\int_0^\infty \sum_{l=0}^K P_k(\hat{S}_1^K \le 0, \ldots, \hat{S}_{n-1}^K \in -\mathrm{d}y, \hat{S}_n^K > 0, \hat{G}_{n-1}^K = l) \\
&\qquad \times \mathrm{d}P_l(\hat{X}_1^K \ge x + y, \hat{G}_1^K = j | \hat{X}_1^K \ge y) \ .
\end{aligned}$$

Further note that, for all $l \ge 0$,

$$\begin{aligned}
&P_l(\hat{X}_1^K > x + y, \hat{G}_1^K = j \,|\, \hat{X}_1^K > y) \\
&\quad = \frac{P(\hat{X}_1^K > x + y, \hat{G}_1^K = j \,|\, \hat{G}_0^K = l)}{P(\hat{X}_1^K > y \,|\, \hat{G}_0^K = l)} \\
&\quad = \frac{\frac{\pi_j^K}{\pi_l^K} p_{jl}^K - \frac{\pi_j^K}{\pi_l^K} P(X_1^K \le x + y, G_1^K = l \,|\, G_0^K = j)}{1 - \sum_{i=0}^K \frac{\pi_i^K}{\pi_l^K} P(X_1^K \le y, G_1^K = l \,|\, G_0^K = i)} \\
&\quad = \frac{\pi_j^K}{\pi_l^K} \frac{p_{jl}^K - P(X_1^K \le x + y, G_1^K = l \,|\, G_0^K = j)}{P(X_1^K > y \,|\, G_1^K = l)} \ .
\end{aligned}$$

On the other hand, for all $j \ge 0$,

$$\begin{aligned}
&P_j(X_1^K \ge x + y, G_1^K = l \,|\, X_1^K \ge y) \\
&\quad = \frac{P(X_1^K \ge x + y, G_1^K = l \,|\, G_0^K = j) P(G_0^K = j)}{P(G_0^K = j, X_1^K \ge y)} \\
&\quad = \frac{p_{jl}^K - P(X_1^K \le x + y, G_1^K = l \,|\, G_0^K = j)}{P(X_1^K > y \,|\, G_0^K = j)} \ ,
\end{aligned}$$

i.e., for all $l \ge 0$,

$$\begin{aligned}
&P_l(\hat{X}_1^K \ge x + y, \hat{G}_1^K = j \,|\, \hat{X}_1^K > y) \\
&\quad = \frac{\pi_j^K}{\pi_l^K} \frac{P(X_1^K \ge y \,|\, G_0^K = j)}{P(X_1^K \ge y \,|\, G_1^K = l)} P_j(X_1^K \ge x + y, G_1^K = l \,|\, X_1^K \ge y) \ .
\end{aligned}$$

Thus by (4.5) and (4.9) for $j > 0$,

$$\begin{aligned}
&\mathrm{d}g_{kj}^{(n)}(x) \\
&\quad = -\int_0^\infty \sum_{l=0}^K P_k(\hat{S}_1^K \le 0, \ldots, \hat{S}_{n-1}^K \in -\mathrm{d}y, \hat{S}_n^K > 0, \hat{G}_{n-1}^K = l)
\end{aligned}$$

$$
\begin{aligned}
&\qquad\qquad \times \mathrm{d}P_l(\hat{X}_1^K \geq x+y, \hat{G}_1^K = j \,|\, \hat{X}_1^K \geq y) \\
= & \; -\int_0^\infty \sum_{l=0}^K \frac{\pi_j^K}{\pi_l^K} \frac{P(X_1^K \geq y \,|\, G_0^K = j)}{P(X_1^K \geq y \,|\, G_1^K = l)} P_k(\hat{S}_1^K \leq 0, \ldots, \hat{S}_{n-1}^K \in -\mathrm{d}y, \hat{S}_n^K > 0, \hat{G}_{n-1}^K = l) \\
&\qquad\qquad \times \mathrm{d}P_j(X_1^K \geq x+y, G_1^K = l \,|\, X_1^K \geq y) \\
= & \; \int_0^\infty \sum_{l=0}^K \frac{\pi_j^K}{\pi_l^K} \frac{P(X_1^K \geq y \,|\, G_0^K = j)}{P(X_1^K \geq y \,|\, G_1^K = l)} P_k(\hat{S}_1^K \leq 0, \ldots, \hat{S}_{n-1}^K \in -\mathrm{d}y, \hat{S}_n^K > 0, \hat{G}_{n-1}^K = l) \\
\times & \; \Big(\sum_{m=0}^{j-1} \sum_{i=0}^{m} a_{jl}^{(m,i)}(y) x^i \mathrm{e}^{-x/c}\Big)\mathrm{d}x = \sum_{i=0}^{j-1} A_{kj}^{(n,i)} x^i \mathrm{e}^{-x/c} \mathrm{d}x \,,
\end{aligned}
$$

for some real matrices $A^{(n,i)}$. Similarly,

$$
\mathrm{d}g_{k0}^{(n)}(x) = A_k^{(0,n)} \mathrm{e}^{-x/c} \mathrm{d}x \,,
$$

for some vectors $A^{(0,n)}$. Hence, for $j > 0$ and $\mathrm{Re}\,\omega \geq 0$ we have

$$
\begin{aligned}
\chi_{kj}^+(\omega) &= \sum_{i=0}^{j-1} \sum_{n=1}^{\infty} A_{kj}^{(n,i)} \int_0^\infty \mathrm{e}^{-\omega x} x^i \mathrm{e}^{-x/c} \mathrm{d}x \\
&= \sum_{i=1}^{j} A_{kj}^{(i)} \Gamma(\omega)^i \,,
\end{aligned}
$$

for some real matrices $A^{(i)}$ $(i = 1, \ldots, K)$, where

$$
\Gamma(\omega) = \frac{1}{1+\omega c} \,.
$$

Thus

$$
\left\{\chi_{kj}^+(\omega)\right\}_{k=0,j=1}^K = \sum_{i=1}^K A^{(i)} \Gamma(\omega)^i \,, \tag{4.10}
$$

where $A^{(i)}$ are real $(K+1) \times K$ matrices fulfilling

$$
A_{kj}^{(i)} = 0 \quad \text{if } j < i. \tag{4.11}
$$

Moreover,

$$
\chi_{k0}^+(\omega) = A_k^{(0)} \Gamma(\omega) \,, \tag{4.12}
$$

for some vector $A^{(0)}$. Though the LST $\chi^+(\omega)$ is defined only for $\mathrm{Re}\,\omega \geq 0$, the r.h.s. of (4.10) and (4.12) is well-defined and analytic whenever

$$
\omega \in \mathbb{C} \quad \text{and} \quad \omega \neq -1/c \,.
$$

We denote this extension also by $\chi^+(\omega)$.

Define the matrix

$$
\Phi(\omega) = \Big\{\int_{-\infty}^{+\infty} \exp(-\omega x)\, \mathrm{d}Q_{ij}^K(x)\Big\}_{i,j=0}^K \,.
$$

According to Presman [23], Arjas [2] and Miller [17, 18] the following Wiener-Hopf factorization

$$\mathbb{I} - \Phi(\omega) = \Pi^{-1}[\mathbb{I} - \chi^+(\omega)]'\Pi[\mathbb{I} - \chi^-(\omega)] \tag{4.13}$$

holds in the region $-1/c < \text{Re } \omega \leq 0$. Under stability condition (4.3), we get from Asmussen [4], Prop. 4.2, and Miller [19], that for $\text{Re } \omega < 0$

$$\text{Spr}(\chi^-(\omega)) < \text{Spr}(\chi^-(0)) = 1 \ ,$$

where $\text{Spr}(A)$ is the spectral radius of the matrix A. Hence

$$\sum_{n=0}^{+\infty} \chi^-(\omega)^n = (\mathbb{I} - \chi^-(\omega))^{-1}$$

exists for all $\text{Re } \omega < 0$. Thus from (4.13), for $-1/c < \text{Re } \omega < 0$,

$$\Pi(\mathbb{I} - \Phi(\omega)) \left[\mathbb{I} - \chi^-(\omega)\right]^{-1} = (\mathbb{I} - \chi^+(\omega))'\Pi$$

and analytically extending the functions on both sides of this identity to the region $\{\omega : \text{Re } \omega < 0, \omega \neq -1/c\}$ we have

$$\det(\mathbb{I} - \Phi(\omega)) = 0 \quad \Longrightarrow \quad \det(\mathbb{I} - \chi^+(\omega)) = 0 \ .$$

Note that

$$\Phi(\omega) = Y(\omega) Z(\omega) \ ,$$

where Z is a diagonal matrix with the following entries

$$\begin{aligned} Z_{00}(\omega) &= \frac{1}{(\omega c + 1)} \ , \\ Z_{ii}(\omega) &= \frac{1}{(\omega c + 1)^i} \ , \ i = 1, \ldots, K \ , \end{aligned}$$

on the diagonal and

$$Y_{ij}(\omega) = p_{ij}^K \int_0^\infty \mathrm{e}^{-\omega x} \ \mathrm{d}F_{ij}^a(x) = p_{ij}^K \left(\frac{\kappa/\beta + \lambda}{\kappa/\beta + \lambda - (1-a)\omega} \right)^{\kappa i + j}$$

for $i > 0$ and

$$Y_{0j}(\omega) = p_{0j}^K \int_0^\infty \mathrm{e}^{-\omega x} \ \mathrm{d}F_{0j}^0(x) = p_{0j}^K \left(\frac{\kappa/\beta + \lambda}{\kappa/\beta + \lambda - \omega} \right)^{\kappa + j} \frac{\lambda}{\lambda - \omega} \ .$$

From Rouché's Theorem for matrices (see de Smit [24]) the number of roots of the equation

$$\det(\mathbb{I} - \Phi(\omega)) = 0 \tag{4.14}$$

in the region $\text{Re } \omega < 0$ is the same as the number of poles of $\det \ \Phi(\omega) = \det Y(\omega) \ \det Z(\omega)$, i.e., since $\det Y(\omega)$ has no poles for $\text{Re } \omega < 0$, it equals the number r of poles of $\det Z(\omega)$. We have

$$r = 1 + \sum_{j=1}^{K} j = 1 + \frac{K(K+1)}{2} \ .$$

Denote the roots of equation (4.14) by $\omega_1, \dots, \omega_r$.

Condition 1.

$\omega_1, \dots, \omega_r$ are all distinct and different from $-1/c$.

As in de Smit and Regterschot [25] we see that this condition will be always satisfied except possibly for countably many values of one of the parameters λ, c, β. In all numerical examples that we analysed, this condition was satisfied. See also de Smit and Regterschot [25] and Çinlar [9] for further comments on this condition. Let h_i be the left eigenvectors with eigenvalue 1 of the matrix $\Phi(\omega_i)$:

$$h_i \Phi(\omega_i) = h_i \qquad \text{for } i = 1, \dots, r.$$

From (4.13) we have

$$h_i[\mathbb{1} - \Phi(\omega_i)] = h_i \Pi^{-1} [\mathbb{1} - \chi^+(\omega_i)]' \Pi [\mathbb{1} - \chi^-(\omega_i)] = 0 \ ,$$

i.e. also

$$h_i \Pi^{-1} [\chi^+(\omega_i)]' = h_i \Pi^{-1}, \qquad i = 1, \dots, r \ , \tag{4.15}$$

where $\chi^+(\omega)$ here means the extension into $\operatorname{Re} \omega < 0$. Taking the transpose on each side of (4.15) yields, with the notation $l_i = \Pi^{-1} h_i'$,

$$\chi^+(\omega_i) l_i = l_i, \qquad i = 1, \dots, r \, . \tag{4.16}$$

Define the $(K+1) \times (1 + K(K+1))$ matrix

$$A = \begin{pmatrix} A^{(0)} & 0 & A^{(1)} & 0 & A^{(2)} & \cdots & 0 & A^{(K)} \end{pmatrix}$$

with $A^{(i)}$ given above and 0 denoting a column vector of zeros, and let

$$\tilde{l}_i = \begin{pmatrix} l_{i,0}\Gamma(\omega_i) & l_i'\Gamma(\omega_i) & l_i'\Gamma^2(\omega_i) & \cdots & l_i'\Gamma^K(\omega_i) \end{pmatrix}' \ ,$$

where $l_{i,0}$ denotes the first element of the vector l_i. Further, define matrices $L = (l_i)$ and $\tilde{L} = (\tilde{l}_i)$. Then, (4.16) is equivalent to

$$A\tilde{L} = L \ . \tag{4.17}$$

Let $\overline{L}$ be a square matrix of dimension r created from those rows of $\tilde{L}$ which are not multiplied by a zero column of matrix A, i.e., $\overline{L} = (\operatorname{row}(\tilde{L}, k))$ such that

$$k \neq 1 + (i-1)(K+1) + j \qquad \text{for } i = 1, \dots, K \text{ and } j = 1, \dots, i.$$

Condition 2.

The rank of matrix $\overline{L}$ is equal to r.

The remark made after Condition 1 also applies here. From (4.11) the number of unknown variables in a row of matrix A is equal to r. Thus by Conditions 1 and 2 the system of equations (4.17) has a unique solution. Summarizing, from (4.7), (4.10) and (4.17) we get the following theorem.

Theorem 4.2 *Under conditions* 1 *and* 2, *the* LST $\mathbb{E}_i(\exp(-\omega W_2^K), G^K = j)$ *is the* j^{th} *entry of the vector*

$$\mathbb{E}_i(\exp(-\omega W_2^K)) = \sum_{j=0}^{K} \mathbb{E}_i(\exp(-\omega W_2^K), G^K = j) = \boldsymbol{e}'\Pi[\mathbb{I} - \chi^+(\omega)]^{-1}[\mathbb{I} - \chi^+(0)]\boldsymbol{e} \ ,$$

where $\chi^+(\omega)$ *is given by* (4.10) *and* (4.12), *and the matrices* $A^{(i)}$ *are obtained by solving* (4.17).

From Theorem 4.2, the mean waiting time $\mathbb{E}W_2^K$ can be obtained. We use the notation

$$\chi_{kj}^+ = c\sum_{i=1}^{K} A_{kj}^{(i)} i \ ,$$

for $j > 0$ and

$$\chi_{k0}^+ = cA_k^{(0)} \ .$$

Corollary 4.1 *The mean waiting time* $\mathbb{E}W_2^K$ *is given by*

$$\mathbb{E}W_2^K = \boldsymbol{e}'\Pi[\mathbb{I} - \chi^+(0)]^{-1}\chi^+\boldsymbol{e} \, ,$$

where

$$\chi_{k0}^+(0) = A_k^{(0)} \, , \qquad \chi_{kj}^+(0) = \sum_{i=1}^{K} A_{kj}^{(i)} \qquad \textit{for } j > 0.$$

Proof. Note that

$$\mathbb{E}W_2^K = -\frac{\mathrm{d}}{\mathrm{d}\omega}\,\mathbb{E}\mathrm{e}^{-\omega W_2^K}\Big|_{\omega=0} = -\frac{\mathrm{d}}{\mathrm{d}\omega}\,\boldsymbol{e}'\Pi[\mathbb{I} - \chi^+(\omega)]^{-1}[\mathbb{I} - \chi^+(0)]\boldsymbol{e}\Big|_{\omega=0} \ .$$

Further, $\frac{\mathrm{d}}{\mathrm{d}\omega}\chi^+(\omega)\Big|_{\omega=0} = -\chi^+$. Moreover,

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}\omega}\mathbb{I} = \frac{\mathrm{d}}{\mathrm{d}\omega}[\mathbb{I} - \chi^+(\omega)][\mathbb{I} - \chi^+(\omega)]^{-1} \\ &= -\frac{\mathrm{d}}{\mathrm{d}\omega}\chi^+(\omega)[\mathbb{I} - \chi^+(\omega)]^{-1} + [\mathbb{I} - \chi^+(\omega)]\,\frac{\mathrm{d}}{\mathrm{d}\omega}[\mathbb{I} - \chi^+(\omega)]^{-1} \, , \end{aligned}$$

and thus

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}\omega}[\mathbb{I} - \chi^+(\omega)]^{-1}\Big|_{\omega=0} = \left([\mathbb{I} - \chi^+(\omega)]^{-1}\,\frac{\mathrm{d}}{\mathrm{d}\omega}\chi^+(\omega)[\mathbb{I} - \chi^+(\omega)]^{-1}\right)\Big|_{\omega=0} \\ &= \ [\mathbb{I} - \chi^+(0)]^{-1}\,\frac{\mathrm{d}}{\mathrm{d}\omega}\chi^+(\omega)\Big|_{\omega=0}[\mathbb{I} - \chi^+(0))]^{-1} = -[\mathbb{I} - \chi^+(0)]^{-1}\chi^+[\mathbb{I} - \chi^+(0))]^{-1} \ . \end{aligned}$$

□

Remark 4.1 Note that all considerations can be further generalized to the case that the distribution of C_k is not necessarily an exponential distribution but has a density f satisfying $f(x+y) = \sum_i h_i(x)g_i(y)$ for some functions h_i and g_i. This is in particular the case for densities that are exponential polynomials, which is, by Borovkov [6], p. 105, equivalent to the assumption that the LST of C_k is rational.

We have used the numerical algorithm described in this section and calculated the mean of the waiting time for a few different parameters of our model. We take $\kappa = 2$ and $\beta = 1$.

	$\lambda = 0,5 \quad c = 0,5$ $a = 0,5$	$\lambda = 0,5 \quad c = 0,5$ $a = 0,3$	$\lambda = 0,3 \quad c = 0,5$ $a = 0,5$	$\lambda = 0,3 \quad c = 0,3$ $a = 0,5$
$K = 3$	$0,1965$	$0,1325$	$0,0997$	$0,0336$
$K = 4$	$0,2070$	$0,1555$	$0,1513$	$0,0805$

Table 1: $\mathbb{E}W_2^K$ for several parameter values

Acknowledgement

The authors are indebted to Dee Denteneer (Philips Research Laboratories) for many useful discussions and suggestions.

References

[1] I.J.B.F. Adan, S.C. Borst and S. Shankar. *A queueing theoretic approach to contention resolution protocol in HFC access networks.* Manuscript in preparation (2002).

[2] E. Arjas. *On a fundamental identity in the theory of semi-Markov processes.* Adv. Appl. Prob. **4** (1972), 258–270.

[3] E. Arjas. *On the use of a fundamental identity in the theory of semi-Markov queues.* Adv. Appl. Prob. **4** (1972), 271–284.

[4] S. Asmussen. *Aspects of matrix Wiener-Hopf factorization in Applied Probability.* Math. Scientist **14** (1989), 101–116.

[5] S. Asmussen. Applied Probablity and Queues. *John Wiley & Sons*, New York 1987.

[6] A.A. Borovkov. Stochastic Processes in Queueing Theory. *Springer-Verlag*, New York 1976.

[7] O.J. Boxma and J.W. Cohen. *The $M/G/1$ queue with permanent customers.* IEEE J. Sel. Areas Commun. **9** (1991), 179–184.

[8] J.I. Capetanakis. *Tree algorithms for packet broadcast channels.* IEEE Trans. Inform. Theory **23** (1979), 505–515.

[9] E. Çinlar. *Queues with semi-Markov arrivals.* J.Appl.Prob. **4** (1967), 365–379.

[10] J.W. Cohen. *On the optimal switching level for an $M/G/1$ queueing system.* Stoch. Proc. Appl. **4** (1976), 297–316.

[11] D. Denteneer and V. Pronk. *On the number of contenders in a contention tree.* Proc. 14$^{\text{th}}$ ITC Specialist Seminar, Girona (2001), 105–112.

[12] P. Flajolet and P. Mathys. *Q-ary collision resolution algorithms in random-access systems with free or blocked channel access.* IEEE Trans. Inform. Theory **31** (1985), 217–243.

[13] A.J.E.M. Janssen and M.J.M. de Jong. *Analysis of contention tree algorithms.* IEEE Trans. Inform. Theory **46** (2000), 2163 – 2172.

[14] F.P. Kelly. *Mathematical modelling of the Internet.* In "Mathematics Unlimited - 2001 and Beyond", *Springer-Verlag*, Berlin 2001, 685 – 702.

[15] J.O. Limb and D. Sala. *A protocol for efficient transfer of data over hybrid fiber/coax systems.* IEEE/ACM Transactions on Networking **5** (1997), 872–881.

[16] R.M. Loynes. *Stationary waiting time distributions for single server queues.* Ann. Math. Statist. **33** (1962), 1323–1339.

[17] H.D. Miller. *A matrix factorization problem in the theory of random variables defined on a finite Markov chain.* Proc. Camb. Philos. Soc. **58** (1962), 268–285.

[18] H.D. Miller. *Absorption probabilities for sums of random variables defined on a finite Markov chain.* Proc. Camb. Philos. Soc. **58** (1962), 286–298.

[19] H.D. Miller. *A convexity property in the theory of random variables defined on a finite Markov chain.* Ann. Math. Statist. **33** (1961), 1260–1270.

[20] M.F. Neuts. *The single server queue with Poisson input and semi-Markov service times.* J. Appl. Prob. **3** (1966), 202–230.

[21] L. Örmeci and J.A.C. Resing. *Two tandem queueing models.* Manuscript, Eindhoven University of Technology, Dept. of Mathematics and Computer Science (2000).

[22] N.U. Prabhu and L.C. Tang. *Markov-modulated single-server queueing systems.* J. Appl. Probab. **31A** (1994), 169–184.

[23] E.L. Presman. *Factorization methods and boundary problems for sums of random variables given on Markov chains.* Math. USSR Izvestija **3** (1969), 815–852.

[24] J.H.A. de Smit. *Explicit Wiener-Hopf factorizations for the analysis of multidimensional queues.* Proc. Internat. Symp. on Semi-Markov processes, ed. J. Janssen, Plenum, New York (1986), 293–309.

[25] J.H.A. de Smit and G.J.K. Regterschot. *A semi-Markov queue with exponential service times.* Proc.Internat.Symp. on Semi-Markov processes, ed. J. Janssen, Plenum, New York (1986), 369–382.

[26] J.H.A. de Smit. *The single server semi-Markov queue.* Stoch. Proc. Appl. **22** (1986), 37–50.

[27] H. Takagi. Queueing Analysis, Vol. 1. North-Holland, Amsterdam 1991.

[28] B.S. Tsybakov and V.A. Mikhailov. *Free synchronous packet access in a broadcast channel with feedback.* Probl. Inform. Trans. **14** (1978), 259–280.