

AN M/G/1 QUEUE WITH CUSTOMER COLLECTION

S.C. Borst¹, O.J. Boxma^{1,2}, M.B. Combé^{1*}

¹CWI - Centre for Mathematics & Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

²Faculty of Economics, Tilburg University
P.O. Box 90153, 5000 LE Tilburg, The Netherlands

Keywords & Phrases: $M/G/1$ queue, dependence, sojourn time, waiting time, queue length, busy period.

ABSTRACT

This paper analyses a variant of the $M/G/1$ queue in which the service times of arriving customers depend on the length of the interval between their arrival and the previous arrival. The dependence structure under consideration arises when individual customers arrive according to a Poisson process, while customer collectors are sent out according to a Poisson process to collect the customers and to bring them to the service facility. In this case collected numbers of customers, and hence total collected service requests, are positively correlated with the corresponding collect intervals. Viewing a batch of collected customers as one (super)customer gives rise to an $M/G/1$ queue with a positive cor-

*This author was supported by NFI.

relation between service times of such customers and their interarrival times. Both for individual customers and for supercustomers we derive the transforms of the sojourn time, waiting time and queue length distributions. We also compare our results with those for the ordinary $M/G/1$ queue without dependence.

1 INTRODUCTION

Consider the following situation. Customers arrive at pick-up points according to independent Poisson processes. At these pick-up points they wait for a bus to bring them to a single-server service facility (e.g., the check-in counter of a hotel). Busses with unlimited customer capacity move according to a fixed route along the pick-up points, with fixed speed, collecting all waiting customers that they encounter and finally delivering all collected customers at the service facility. The intervals between the starts of successive bus tours are exponentially distributed. Because of the fixed tour length, the arrival process of busses at the service facility is a Poisson process. Viewing the batch of customers brought to the service facility by a bus as one supercustomer, the service facility very closely resembles an ordinary $M/G/1$ queue; the only difference is that *the service time of a supercustomer depends on the previous interarrival time*. Indeed, if two consecutive busses arrive at a relatively long (short) interval, then the second bus is likely to pick up relatively many (few) customers: the interarrival time and the size of the picked-up batch are positively correlated, and hence so are the interarrival time and the supercustomer's service time.

In the present paper we analyse the $M/G/1$ queue with the above-sketched structure of the correlation between interarrival and service times. Our motivation for this analysis is twofold. Firstly, in the vast literature on single server queues it is almost exclusively assumed that there is no dependence between arrival intervals, between service times and between interarrival and service times, although dependencies between these quantities occur in a very natural way. The main reason for ignoring these dependencies seems the mathematical complexity that they almost invariably give rise to. However, the present model allows a very detailed analysis of most performance measures of interest, thus giving valuable insight into the effect that dependencies between interarrival and service times may have on those performance measures. Our second

motive for the analysis is that the correlation structure under consideration seems quite natural in many situations in which customers are collectively brought to a central service facility. Examples are mail pick-up and the pick-up of customers at airport terminals. In computer-communications, one might think of the collection of packets in a 'train' in a Local Area Network with interconnected rings, to be delivered at a bridge queue. Furthermore, modern reservation protocols for the use of transmission slots in high-speed Local and Metropolitan Area Networks also may give rise to customer collection. In particular, in a recent performance evaluation by Boxma, Levy and Yechiali [1992] of the Cyclic-Reservation Multiple-Access (CRMA) protocol a quite similar customer collection procedure occurs. In Boxma, Levy and Yechiali [1992] it is assumed that the collection of customers takes place at fixed intervals, thus giving rise to a $D/G/1$ queue - in which obviously the interarrival and service times are *not* dependent. A more detailed model of the CRMA protocol, taking its backpressure mechanism into account (Nassehi [1989]) would lead to a model that is similar to ours, but in which the arrival process is closer to a deterministic process than to a Poisson process. It would be worthwhile to combine the known $D/G/1$ results and the results from the present paper, to study the performance of CRMA with backpressure. Performance measures like the sojourn time distributions might be approximated by a weighted sum of these distributions for the present $M/G/1$ case and the $D/G/1$ case with weight factors the squared coefficient of variation of the interarrival times and one minus this coefficient.

Model description

Individual customers require service from a service facility with a single server. Their service times are independent, identically distributed stochastic variables with distribution $B(\cdot)$, with mean β_1 , second moment β_2 and Laplace-Stieltjes Transform (LST) $\beta(\cdot)$; $B(0+) = 0$. These individual customers arrive at a pick-up point according to a Poisson process with rate λ . They are collected by a collector and delivered in batches at the service facility at times t_1, t_2, \dots . The collect intervals $\sigma_i = t_i - t_{i-1}$, $i = 1, 2, \dots$, with $t_0 = 0$, are independent, negative exponentially distributed stochastic variables with mean $1/\gamma$. A delivered batch of customers can be viewed as one supercustomer. Supercustomers

apparently arrive at the service facility according to a Poisson process with rate γ . In the sequel the travel time of the collector from the pick-up point to the service facility is assumed to be zero. Thus the arrival of the supercustomer at the service facility coincides with the arrival of the collector at the pick-up point. A non-zero travel time can easily be implemented into our model, by just adding it to the waiting times and sojourn times of individual customers.

Define the total offered traffic load as $\rho := \lambda\beta_1$. In a more general framework Loynes [1962] showed that $\rho < 1$ is a necessary and sufficient condition for the waiting times to have a proper limiting distribution that is independent of the initial conditions. Without proof we claim that in our model the same holds for the other quantities under consideration: sojourn times, queue lengths and busy periods.

After this global model description we consider batch sizes, service times of supercustomers and the correlation structure of the model in some more detail. The number of individual customers, \mathbf{K}_i , constituting the i -th supercustomer generated in a collect interval σ_i of length u has conditional distribution:

$$Pr\{\mathbf{K}_i = n | \sigma_i = u\} = e^{-\lambda u} \frac{(\lambda u)^n}{n!}, \quad u > 0, n = 0, 1, \dots,$$

so

$$\begin{aligned} Pr\{\mathbf{K}_i = n\} &= \int_{u=0}^{\infty} \gamma e^{-\gamma u} Pr\{\mathbf{K}_i = n | \sigma_i = u\} du \\ &= \frac{\gamma}{\gamma + \lambda} \left(\frac{\lambda}{\gamma + \lambda} \right)^n, \quad n = 0, 1, \dots \end{aligned}$$

The service time τ_i of the i -th supercustomer generated in a collect interval σ_i of length u has conditional distribution

$$\begin{aligned} Pr\{\tau_i < t | \sigma_i = u\} &= \sum_{n=0}^{\infty} Pr\{\mathbf{K}_i = n | \sigma_i = u\} B^{n*}(t) \\ &= \sum_{n=0}^{\infty} e^{-\lambda u} \frac{(\lambda u)^n}{n!} B^{n*}(t), \quad t \geq 0, u > 0. \end{aligned} \quad (1.1)$$

Hence

$$E(e^{-\omega \tau_i} | \sigma_i = u) = e^{-\lambda(1-\beta(\omega))u}, \quad Re \omega \geq 0, u > 0.$$

It follows that

$$\begin{aligned} Pr\{\tau_i < t\} &= \sum_{n=0}^{\infty} \frac{\gamma}{\gamma + \lambda} \left(\frac{\lambda}{\gamma + \lambda} \right)^n B^{n*}(t), \quad t \geq 0, \\ E(e^{-\omega\tau_i}) &= \frac{\gamma}{\gamma + \lambda(1 - \beta(\omega))}, \quad Re \omega \geq 0. \end{aligned} \quad (1.2)$$

The service times τ_1, τ_2, \dots of supercustomers are independent, identically distributed stochastic variables. It should be noted that a batch may be empty, and hence a supercustomer may have zero service time:

$$\begin{aligned} Pr\{\mathbf{K}_i = 0 | \sigma_i = u\} &= Pr\{\tau_i = 0 | \sigma_i = u\} = e^{-\lambda u}, \quad u > 0, \\ Pr\{\mathbf{K}_i = 0\} &= Pr\{\tau_i = 0\} = \frac{\gamma}{\gamma + \lambda}. \end{aligned} \quad (1.3)$$

From (1.2) it follows that

$$E\tau_i = \frac{\lambda\beta_1}{\gamma}, \quad E\tau_i^2 = \frac{\lambda\beta_2}{\gamma} + 2 \left(\frac{\lambda\beta_1}{\gamma} \right)^2. \quad (1.4)$$

The bivariate LST of σ_i and τ_i follows from (1.1):

$$E(e^{-\zeta\sigma_i - \omega\tau_i}) = \frac{\gamma}{\gamma + \zeta + \lambda(1 - \beta(\omega))}, \quad Re \zeta \geq 0, Re \omega \geq 0,$$

yielding

$$\begin{aligned} Cov(\sigma_i, \tau_i) &= \frac{\lambda\beta_1}{\gamma^2}, \\ correl(\sigma_i, \tau_i) &= \left[1 + \frac{\gamma\beta_2}{\lambda\beta_1^2} \right]^{-1/2} \geq 0. \end{aligned}$$

Note that $correl(\sigma_i, \tau_i) \rightarrow 1$ for $\gamma \downarrow 0$ and also for $\lambda \rightarrow \infty$, whereas $correl(\sigma_i, \tau_i) \rightarrow 0$ for $\gamma \rightarrow \infty$ and for $\lambda \downarrow 0$. For $\gamma \rightarrow \infty$ the queue approaches an ordinary M/G/1 queue, as individual customers are collected instantaneously.

Remark 1.1 Instead of assuming that there is a single Poisson arrival stream of individual customers with service time distribution $B(\cdot)$, we could also have allowed several independent Poisson arrival streams at various pick-up points, with different service time distributions. A distinction between the various

arrival streams may be useful in certain applications, where e.g. waiting times of individual customers must be determined taking into account the location of the pick-up point and the travel time of the collector to the service facility. This refinement can easily be implemented into our model, without seriously complicating the analysis. □

Related literature

Very few studies have appeared that analyse a queueing system with correlation between the interarrival and service times. Cidon et al. [1991a] consider a correlation between the service time of a customer and the *subsequent* interarrival time; the waiting time for such a correlated queue can be analysed by studying the recurrence relation for the waiting time,

$$\mathbf{W}_{i+1} = \max(0, \mathbf{W}_i + \tau_i - \sigma_{i+1}),$$

in an ordinary $GI/GI/1$ queue with similarly distributed $\tau_i - \sigma_{i+1}$ and with \mathbf{W}_i and $\tau_i - \sigma_{i+1}$ independent.

Like the present paper, a number of papers has been devoted to the single server queue with correlation between the service time of a customer and the *preceding* interarrival time. Conolly [1968] and Conolly and Hadidi [1969] consider an $M/M/1$ queue in which the service time and the preceding interarrival time are linearly related. Conolly and Choo [1979] study an $M/M/1$ queue in which σ_i and τ_i have a bivariate exponential distribution with density

$$g(s, t) = \lambda\mu(1 - r)e^{-\lambda s - \mu t} I_0[2\{\lambda\mu r s t\}^{1/2}], \quad (1.5)$$

where $I_0[z]$ is a zero order modified Bessel function of the first kind, and where $r \in [0, 1)$ is the correlation between σ_i and τ_i . For $r = 0$ the queue reduces to an ordinary $M/M/1$ queue. The marginal distributions of σ_i and τ_i are negative exponential. Conolly and Choo analyse the waiting time distribution for this correlated $M/M/1$ queue, showing that its density can be expanded in a series of partial fraction terms. Their numerical calculations reveal that the positive correlation leads to a considerable reduction in mean waiting time. For the same model, (i) Hadidi [1981] shows that the waiting times are hyperexponentially distributed; (ii) Hadidi [1985] examines the sensitivity of the waiting time distribution to the value of the correlation coefficient; (iii) Lan-

garis [1987] studies the busy period distribution. However, the starting point of the latter study seems to be wrong: it is assumed that a customer who starts a general busy period has an ordinary service time, whereas in this correlated queue a customer who starts a busy period is likely to have a relatively long interarrival time and hence a relatively long service time. This flaw is reported in more detail in Borst & Combé [1992].

The paper that is closest related to the present study is Takahashi [1971]. He uses the terminology of ‘gates’ instead of ‘busses’: Arriving customers first join a queue at the gate of the system; at exponentially distributed intervals the gate opens, and the customers at the first queue move to the second queue, where r servers are available. The gate closes immediately after all customers in the first queue have moved to the second queue. Takahashi’s study is more restrictive than ours in the sense that he only allows exponentially distributed service times, and that he only studies sojourn times and queue lengths of *individual* customers. His study is more general than ours in the sense that he allows multiple servers. Furthermore, he also briefly considers the case in which the gate opens at fixed intervals.

Several other studies have been devoted to queueing systems with two stages of waiting, with a gate at the first queue and service provided only at the second queue; see e.g. Coleman [1973] and Ali & Neuts [1984]. However, in these studies the opening and closing of the gate is determined by the queue lengths, rather than by a Poisson point process.

Organization of the paper

In Section 2 (3) we derive the LST of the sojourn (waiting) time distributions of supercustomers as well as of individual customers. The mean waiting time of supercustomers, EW , is compared with the mean waiting times in (i) the $M/G/1$ queue without collection, in which individual customers do not wait to be picked up but immediately join one common queue for service ($EW_{M/G/1}$), and (ii) the $M/G/1$ queue of supercustomers in the case of *independence* of interarrival and service times (EW_I). It is proven that $EW_I \geq EW \geq EW_{M/G/1}$.

In Section 4 we obtain the joint distribution of the numbers of those individual customers waiting to be collected and those already collected but waiting for service, immediately after the departure of an individual customer. The

queue length distribution of supercustomers is also derived. Section 5 is devoted to a discussion concerning the busy period distribution in the $M/G/1$ queue with dependence. The mean busy period is easily obtained; the busy period *distribution* gives rise to mathematical difficulties which are briefly discussed but not solved. Section 6 contains some numerical results, exposing the influence of the dependence on waiting time means and busy period variance. Conclusions and suggestions for further research are presented in Section 7.

2 THE SOJOURN TIME

Denote by \mathbf{R}_i the sojourn time of the i -th supercustomer, $i = 1, 2, \dots$, i.e. the time from the arrival of the supercustomer until the completion of the service of the last individual customer belonging to the supercustomer. Let $r_i(\omega) := E(e^{-\omega \mathbf{R}_i})$ for $Re \omega \geq 0$, $i = 1, 2, \dots$. Starting from the recurrence relation

$$\mathbf{R}_i = \max(0, \mathbf{R}_{i-1} - \sigma_i) + \tau_i, \quad i = 2, 3, \dots, \quad (2.1)$$

a straightforward calculation yields

$$r_i(\omega) = \frac{\gamma}{\gamma + \lambda(1 - \beta(\omega))} r_{i-1}(\gamma + \lambda(1 - \beta(\omega))) + \frac{\gamma}{\gamma - \omega + \lambda(1 - \beta(\omega))} [r_{i-1}(\omega) - r_{i-1}(\gamma + \lambda(1 - \beta(\omega)))]. \quad (2.2)$$

A detailed derivation of (2.2) is given in Borst et al. [1992].

As observed in the introduction, for $\rho < 1$ the sojourn times \mathbf{R}_i have a proper limiting distribution, $R(\cdot)$, for $i \rightarrow \infty$. Denote by \mathbf{R} a stochastic variable with this distribution. Let $r(\omega) := E(e^{-\omega \mathbf{R}})$ for $Re \omega \geq 0$. From (2.2) we obtain

$$r(\omega) = \frac{\gamma \omega}{\omega - \lambda(1 - \beta(\omega))} \frac{r(\gamma + \lambda(1 - \beta(\omega)))}{\gamma + \lambda(1 - \beta(\omega))}, \quad Re \omega \geq 0. \quad (2.3)$$

We solve the above functional equation for $r(\cdot)$ below. We first derive some preliminary results from (2.3).

Remark 2.1 Letting $\omega \rightarrow \infty$ in (2.3) we obtain

$$Pr\{\mathbf{R} = 0\} = \frac{\gamma}{\gamma + \lambda} r(\gamma + \lambda). \quad (2.4)$$

This formula may also be obtained directly from (2.1) and (1.3):

$$Pr\{\mathbf{R}_i = 0\} = \int_{u=0}^{\infty} \gamma e^{-\gamma u} Pr\{\tau_i = 0, \mathbf{R}_{i-1} < \sigma_i \mid \sigma_i = u\} du =$$

$$\int_{u=0}^{\infty} \gamma e^{-\gamma u} e^{-\lambda u} Pr\{\mathbf{R}_{i-1} < u\} du = \frac{\gamma}{\gamma + \lambda} r_{i-1}(\gamma + \lambda).$$

□

Remark 2.2 Letting $\omega \downarrow 0$ in (2.3) we obtain

$$r(\gamma) = 1 - \lambda\beta_1. \tag{2.5}$$

Again, this formula may also be obtained directly:

$$r_{i-1}(\gamma) = \int_{t=0}^{\infty} e^{-\gamma t} dPr\{\mathbf{R}_{i-1} < t\} = Pr\{\mathbf{R}_{i-1} < \sigma_i\}.$$

The latter term equals the probability that an arriving supercustomer sees the server idle, while obviously the probability that the server is idle equals $1 - \lambda\beta_1$. Because of the PASTA property both probabilities are equal.

□

We now solve the functional equation (2.3) for $r(\cdot)$. Define

$$f(\omega) := \frac{\gamma\omega}{\omega - \lambda(1 - \beta(\omega))}, \quad Re\omega \geq 0, \tag{2.6}$$

$$g(\omega) := \gamma + \lambda(1 - \beta(\omega)), \quad Re\omega \geq 0. \tag{2.7}$$

Then we can write

$$r(\omega) = \frac{f(\omega)}{g(\omega)} r(g(\omega)). \tag{2.8}$$

Let

$$g^{(0)}(\omega) := \omega, \quad Re\omega \geq 0,$$

$$g^{(h)}(\omega) := g(g^{(h-1)}(\omega)), \quad Re\omega \geq 0, \quad h = 1, 2, \dots$$

Iterating (2.8) we find

$$r(\omega) = r(g^{(M+1)}(\omega)) \prod_{h=0}^M \frac{f(g^{(h)}(\omega))}{g^{(h+1)}(\omega)}, \quad Re\omega \geq 0,$$

for any non-negative integer M .

Lemma 2.1

- (i). The equation $\omega = g(\omega)$, $Re\omega \geq 0$ has a unique solution ω^* . ω^* is real.
- (ii). $\lim_{M \rightarrow \infty} g^{(M)}(\omega) = \omega^*$ for all ω with $Re\omega \geq 0$.
- (iii). $\prod_{h=0}^{\infty} \frac{f(g^{(h)}(\omega))}{g^{(h+1)}(\omega)}$ converges for all ω with $Re\omega \geq 0$.

Proof

(i) follows from Rouché's theorem; (ii) and (iii) are based on the observation that $|g^{(h+1)}(\omega) - g^{(h)}(\omega)| \leq \rho |g^{(h)}(\omega) - g^{(h-1)}(\omega)|$ for all ω with $Re\omega \geq 0$. A detailed proof is given in Appendix A of Borst et al. [1991].

□

Lemma 2.1 implies that

$$r(\omega) = r(\omega^*) \prod_{h=0}^{\infty} \frac{f(g^{(h)}(\omega))}{g^{(h+1)}(\omega)}, \quad Re\omega \geq 0. \quad (2.9)$$

Putting $\omega = 0$ in (2.9) we find

$$r(\omega^*) = 1 / \prod_{h=0}^{\infty} \frac{f(g^{(h)}(0))}{g^{(h+1)}(0)},$$

which leads to our main result:

Theorem 2.1

The LST of the sojourn time distribution of a supercustomer is

$$r(\omega) = \prod_{h=0}^{\infty} \frac{f(g^{(h)}(\omega)) / f(g^{(h)}(0))}{g^{(h+1)}(\omega) / g^{(h+1)}(0)}, \quad Re\omega \geq 0, \quad (2.10)$$

with $f(\cdot)$ and $g(\cdot)$ given by (2.6) and (2.7).

The LST in (2.10) could be numerically inverted using a procedure outlined in Abate and Whitt [1992]. Differentiating (2.10),

$$ER = \sum_{h=0}^{\infty} \frac{g^{(h)'}(0)[f(g^{(h)}(0))g'(g^{(h)}(0)) - f'(g^{(h)}(0))g(g^{(h)}(0))]}{f(g^{(h)}(0))g(g^{(h)}(0))}, \quad (2.11)$$

which is used for numerical calculations in Section 6.

We now clarify the meaning of the terms composing (2.3), thus relating \mathbf{R} to the waiting time in an ordinary $M/G/1$ queue. Denote by $\mathbf{W}_{M/G/1}$ a stochastic variable with distribution the stationary distribution of the waiting time in an ordinary $M/G/1$ queue with arrival rate λ and service time distribution function $B(\cdot)$. In the sequel we refer to this queue as the corresponding $M/G/1$ queue without collection. From Cohen [1982] p. 255, we have:

$$E(e^{-\omega \mathbf{W}_{M/G/1}}) = \frac{(1 - \lambda\beta_1)\omega}{\omega - \lambda(1 - \beta(\omega))}, \quad \text{Re } \omega \geq 0.$$

Denote by \mathbf{H} the sojourn time of a supercustomer leaving no supercustomers behind. Such a sojourn time has distribution function $H(\cdot)$ with

$$dH(t) = \frac{e^{-\gamma t} dR(t)}{\int_{u=0}^{\infty} e^{-\gamma u} dR(u)}, \quad t > 0,$$

and, cf. (2.4)

$$H(0+) = \frac{\gamma}{\gamma + \lambda} \frac{r(\gamma + \lambda)}{r(\gamma)}.$$

Denote by \mathbf{U} the amount of work arriving during such a sojourn time.

$$E(e^{-\omega \mathbf{U}}) = \frac{\int_{t=0}^{\infty} e^{-\lambda(1-\beta(\omega))t} e^{-\gamma t} dR(t)}{\int_{u=0}^{\infty} e^{-\gamma u} dR(u)} = \frac{r(\gamma + \lambda(1 - \beta(\omega)))}{r(\gamma)}, \quad \text{Re } \omega \geq 0. \quad (2.12)$$

So we can write, using (2.5)

$$r(\omega) = E(e^{-\omega \mathbf{W}_{M/G/1}}) E(e^{-\omega \boldsymbol{\tau}}) E(e^{-\omega \mathbf{U}}), \quad \text{Re } \omega \geq 0. \quad (2.13)$$

From (2.13), using (2.5) and (2.12),

$$\begin{aligned} E\mathbf{R} &= E\mathbf{W}_{M/G/1} + E\boldsymbol{\tau} + E\mathbf{U} \\ &= E\mathbf{W}_{M/G/1} + E\boldsymbol{\tau} + \frac{\lambda\beta_1}{1 - \lambda\beta_1} \int_{t=0}^{\infty} t e^{-\gamma t} dR(t), \end{aligned} \quad (2.14)$$

and

$$\begin{aligned} \text{Var}(\mathbf{R}) = & \text{Var}(\mathbf{W}_{M/G/1}) + \text{Var}(\tau) + \frac{(\lambda\beta_1)^2}{1 - \lambda\beta_1} \int_{t=0}^{\infty} t^2 e^{-\gamma t} dR(t) \\ & + \frac{\lambda\beta_2}{1 - \lambda\beta_1} \int_{t=0}^{\infty} t e^{-\gamma t} dR(t) - \left(\frac{\lambda\beta_1}{1 - \lambda\beta_1} \int_{t=0}^{\infty} t e^{-\gamma t} dR(t) \right)^2. \end{aligned}$$

Remark 2.3 Using (2.6) and (2.7) we find that the factor for $h = 0$ of the infinite product (2.10) equals $E(e^{-\omega \mathbf{W}_{M/G/1}})E(e^{-\omega \tau})$. So the remainder of the infinite product equals $\frac{r(\gamma + \lambda(1 - \beta(\omega)))}{r(\gamma)}$. Similarly we find that the term for $h = 0$ of the infinite sum (2.11) equals $E\mathbf{W}_{M/G/1} + E\tau$. So, using (2.5),

the remainder of the infinite sum equals $\frac{\lambda\beta_1}{1 - \lambda\beta_1} \int_{t=0}^{\infty} t e^{-\gamma t} dR(t)$. □

To provide additional insight, we now give a more intuitive derivation of (2.13).

The sojourn time of a supercustomer consists of two phases, viz.:

- (i). its waiting time, i.e. the time needed to do the work associated with the individual customers present at the server upon the supercustomer's arrival;
- (ii). its service time, i.e. the time needed to do the work associated with the individual customers present at the bus stop upon the supercustomer's arrival. (Remember that the arrival of the supercustomer at the server coincides with the arrival of the collector at the bus stop.)

So the sojourn time of a supercustomer equals the amount of work associated with the individual customers present upon its arrival (at the bus stop as well as at the server). Denote by \mathbf{V} the steady state amount of work associated with the individual customers (at the bus stop as well as at the server). Because of the PASTA property ($\stackrel{d}{=}$ denoting equality in distribution)

$$\mathbf{R} \stackrel{d}{=} \mathbf{V}. \tag{2.15}$$

Denote by $\mathbf{V}_{M/G/1}$ a stochastic variable with distribution the stationary distribution of the amount of work in the corresponding $M/G/1$ queue without collection. Denote by \mathbf{Y} a stochastic variable, independent of $\mathbf{V}_{M/G/1}$, with distribution the stationary distribution of the amount of work associated with the individual customers at an arbitrary epoch in a non-serving interval, i.e.

the amount of work associated with the individual customers present at the bus stop when the server is idle. Now the following work decomposition property holds, cf. Boxma [1989]:

$$\mathbf{V} \stackrel{d}{=} \mathbf{V}_{M/G/1} + \mathbf{Y}. \tag{2.16}$$

Because of the PASTA property

$$\mathbf{V}_{M/G/1} \stackrel{d}{=} \mathbf{W}_{M/G/1}. \tag{2.17}$$

The amount of work associated with individual customers at an arbitrary epoch in a non-serving interval, \mathbf{Y} , consists of two components, viz.:

(i). the amount of work associated with individual customers that have arrived during the sojourn time of the last supercustomer (possibly empty), $\mathbf{Y}^{(i)}$. This sojourn time has distribution function $H(\cdot)$. So

$$\mathbf{Y}^{(i)} \stackrel{d}{=} \mathbf{U}. \tag{2.18}$$

(ii). the amount of work associated with individual customers that have arrived during the past non-serving period since the departure of the last supercustomer (possibly empty), $\mathbf{Y}^{(ii)}$. This past non-serving period is negative exponentially distributed with parameter γ , since the non-serving period is a (residual) collect interval. So

$$\mathbf{Y}^{(ii)} \stackrel{d}{=} \tau. \tag{2.19}$$

Moreover, $\mathbf{Y}^{(i)}$ and $\mathbf{Y}^{(ii)}$ are independent, since the individual customers arrive according to a Poisson process.

Combining (2.15) - (2.19) yields (2.13).

We finally study the sojourn time of an *individual* customer, $\tilde{\mathbf{R}}$. Let $\tilde{r}(\omega) := E(e^{-\omega\tilde{\mathbf{R}}})$ for $Re\omega \geq 0$. First we study the number of individual customers, $\tilde{\mathbf{N}}$. We will find $\tilde{r}(\omega)$ from $\tilde{r}(\lambda(1-z)) = E(z^{\tilde{\mathbf{N}}})$, $|z| \leq 1$. Denote by $\mathbf{N}_{M/G/1}$ a stochastic variable with distribution the stationary distribution of the number of customers in the corresponding $M/G/1$ queue without collection. From Cohen [1982] p. 247, we have:

$$E(z^{\mathbf{N}_{M/G/1}}) = \frac{(1-\lambda\beta_1)(1-z)\beta(\lambda(1-z))}{\beta(\lambda(1-z))-z}, \quad |z| \leq 1. \tag{2.20}$$

Denote by $\mathbf{X}^{(i)}$ the number of individual customers present at an arbitrary epoch in a non-serving interval that have arrived during the sojourn time of the last supercustomer (possibly empty). This sojourn time has distribution function $H(\cdot)$. So

$$E(z^{\mathbf{X}^{(i)}}) = \frac{r(\gamma + \lambda(1 - z))}{r(\gamma)}, \quad |z| \leq 1. \quad (2.21)$$

Denote by $\mathbf{X}^{(ii)}$ the number of individual customers present at an arbitrary epoch in a non-serving interval that have arrived during the past non-serving period since the departure of the last supercustomer (possibly empty). This past non-serving period is negative exponentially distributed with parameter γ , since the non-serving period is a (residual) collect interval. So

$$E(z^{\mathbf{X}^{(ii)}}) = \frac{\gamma}{\gamma + \lambda(1 - z)}, \quad |z| \leq 1. \quad (2.22)$$

Observe that $\mathbf{X}^{(i)}$ and $\mathbf{X}^{(ii)}$ are independent, since the individual customers arrive according to a Poisson process and the non-serving period is a (residual) collect interval, not depending on the sojourn time of the last supercustomer. Now the following queue length decomposition holds, cf. Fuhrmann & Cooper [1985]:

$$E(z^{\tilde{\mathbf{N}}}) = E(z^{\mathbf{N}_{M/G/1}})E(z^{\mathbf{X}^{(i)}})E(z^{\mathbf{X}^{(ii)}}), \quad |z| \leq 1. \quad (2.23)$$

Denote by $\mathbf{R}_{M/G/1}$ a stochastic variable with distribution the stationary distribution of the sojourn time in the corresponding $M/G/1$ queue without collection. Substituting $\omega = \lambda(1 - z)$ in (2.23) leads to:

Theorem 2.2

The LST of the sojourn time distribution of an individual customer is

$$\begin{aligned} \tilde{r}(\omega) &= E(e^{-\omega \mathbf{R}_{M/G/1}})E(e^{-\omega \boldsymbol{\sigma}})E(e^{-\omega \mathbf{H}}) \\ &= E(e^{-\omega \mathbf{W}_{M/G/1}})\beta(\omega)\frac{\gamma}{\gamma + \omega}\frac{r(\gamma + \omega)}{r(\gamma)}, \quad \operatorname{Re} \omega \geq 0, \end{aligned} \quad (2.24)$$

with $r(\cdot)$ given by Theorem 2.1.

Remark 2.4 As mentioned in the introduction, Takahashi [1971] studies sojourn times and queue lengths of *individual* customers in the same model as

the present paper, for the special case of negative exponentially distributed service times. For that case, formula (2.24) reduces to formula (2.29) of Takahashi [1971]. To verify this, note that $E(e^{-\omega \mathbf{W}_{M/G/1}})\beta(\omega)\frac{\gamma}{\gamma + \omega}$ equals $g_0(1 - \omega/\lambda)/g_0(1)$ in Takahashi [1971], and that the n -th term in the infinite product (2.10) matches the term $g_{n+1}(1 - \omega/\lambda)/g_{n+1}(1)$ in the infinite product (2.19) of Takahashi [1971].

□

From (2.24), using (2.5),

$$\begin{aligned} E\tilde{\mathbf{R}} &= E\mathbf{W}_{M/G/1} + \beta_1 + E\sigma + E\mathbf{H} \\ &= E\mathbf{W}_{M/G/1} + \beta_1 + \frac{1}{\gamma} + \frac{1}{1 - \lambda\beta_1} \int_0^\infty te^{-\gamma t} dR(t), \end{aligned} \tag{2.25}$$

and

$$\begin{aligned} Var(\tilde{\mathbf{R}}) &= Var(\mathbf{W}_{M/G/1}) + \beta_2 - \beta_1^2 + \frac{1}{\gamma^2} + \\ &\quad \frac{1}{1 - \lambda\beta_1} \int_0^\infty t^2 e^{-\gamma t} dR(t) - \left(\frac{1}{1 - \lambda\beta_1} \int_0^\infty te^{-\gamma t} dR(t) \right)^2. \end{aligned}$$

3 THE WAITING TIME

Denote by \mathbf{W}_i the waiting time of the i -th supercustomer, $i = 1, 2, \dots$, i.e. the time from the arrival of the supercustomer until the start of the service of the first individual customer belonging to the supercustomer. Let $w_i(\omega) := E(e^{-\omega \mathbf{W}_i})$ for $Re \omega \geq 0$, $i = 1, 2, \dots$. Starting from the recurrence relation

$$\mathbf{W}_i = \max(0, \mathbf{R}_{i-1} - \sigma_i), \quad i = 2, 3, \dots, \tag{3.1}$$

a straightforward calculation yields

$$w_i(\omega) = \frac{\gamma r_{i-1}(\omega) - \omega r_{i-1}(\gamma)}{\gamma - \omega}, \quad Re \omega \geq 0. \tag{3.2}$$

As observed in the introduction, for $\rho < 1$ the waiting times \mathbf{W}_i have a proper limiting distribution for $i \rightarrow \infty$. Denote by \mathbf{W} a stochastic variable with this distribution. Let $w(\omega) := E(e^{-\omega \mathbf{W}})$ for $Re \omega \geq 0$. From (3.2) we obtain

$$w(\omega) = \frac{\gamma r(\omega) - \omega r(\gamma)}{\gamma - \omega}, \quad \operatorname{Re} \omega \geq 0. \quad (3.3)$$

From (3.3) or immediately from (3.1), using (2.5),

$$\operatorname{Pr}\{\mathbf{W} = 0\} = 1 - \lambda\beta_1,$$

a result which may also be obtained by applying the PASTA property.

From (3.3), using (1.4) and (2.5),

$$E\mathbf{W} = E\mathbf{R} - E\tau, \quad (3.4)$$

as should be the case since $\mathbf{R}_i = \mathbf{W}_i + \tau_i$ for $i = 1, 2, \dots$. Combining (2.14) and (3.4),

$$E\mathbf{W} = E\mathbf{W}_{M/G/1} + \frac{\lambda\beta_1}{1 - \lambda\beta_1} \int_{t=0}^{\infty} te^{-\gamma t} dR(t). \quad (3.5)$$

Bounds for the integral occurring in (3.5), which are presented in Borst et al. [1991], imply that $\lim_{\gamma \rightarrow \infty} E\mathbf{W} = E\mathbf{W}_{M/G/1}$ whereas $\lim_{\gamma \downarrow 0} E\mathbf{W} = \infty$.

We now compare the mean waiting time of a supercustomer with the mean waiting time in an ordinary $M/G/1$ queue with identical traffic characteristics, but without dependence. Denote by \mathbf{W}_I the waiting time in an ordinary $M/G/1$ queue with arrival rate γ and service time distribution having LST $\frac{\gamma}{\gamma + \lambda(1 - \beta(\omega))}$, cf. (1.2). In the sequel we refer to this queue as the corresponding $M/G/1$ queue without dependence. Using (1.4),

$$E\mathbf{W}_I = \frac{\gamma \left(\frac{\lambda\beta_2}{\gamma} + 2 \left(\frac{\lambda\beta_1}{\gamma} \right)^2 \right)}{2(1 - \lambda\beta_1)} = E\mathbf{W}_{M/G/1} + \frac{\lambda^2\beta_1^2}{\gamma(1 - \lambda\beta_1)}. \quad (3.6)$$

Combining (3.5) and (3.6),

$$E\mathbf{W} = E\mathbf{W}_I + \frac{\lambda\beta_1}{1 - \lambda\beta_1} \left[\int_{t=0}^{\infty} te^{-\gamma t} dR(t) - \frac{\lambda\beta_1}{\gamma} \right]. \quad (3.7)$$

It follows from (3.7) that $E\mathbf{W} \leq E\mathbf{W}_I$: since $r(\cdot)$ is a convex function, using (2.5),

$$\int_{t=0}^{\infty} te^{-\gamma t} dR(t) = -r'(\gamma) \leq \frac{r(0) - r(\gamma)}{\gamma} = \frac{\lambda\beta_1}{\gamma}.$$

The following argument explains why $E\mathbf{W} \leq E\mathbf{W}_I$. From Wolff [1989] p. 279 we have the following formula for the mean total amount of work in the system:

$$E\mathbf{V} = \gamma E[\mathbf{W}\tau] + \gamma E[\tau^2]/2, \tag{3.8}$$

which is an application of the generalized form of Little's law. Using (1.4), (2.14), (2.15), (3.6) and (3.8),

$$E\mathbf{W} = E\mathbf{W}_I + \frac{\gamma}{1 - \lambda\beta_1} Cov(\mathbf{W}, \tau).$$

A supercustomer having a relatively short/long interarrival time is likely to have a relatively long/short waiting time, but also a relatively short/long service time, due to the dependence. So $Cov(\mathbf{W}, \tau) < 0$.

We finally study the waiting time of an *individual* customer. Denote by $\tilde{\mathbf{W}}$ the waiting time of an individual customer. Let $\tilde{w}(\omega) := E(e^{-\omega\tilde{\mathbf{W}}})$ for $Re\omega \geq 0$. From (2.24),

$$\tilde{w}(\omega) = E(e^{-\omega\mathbf{W}_{M/G/1}}) \frac{\gamma}{\gamma + \omega} \frac{r(\gamma + \omega)}{r(\gamma)}, \quad Re\omega \geq 0, \tag{3.9}$$

since the waiting time and the subsequent service time of an *individual* customer are independent. From (3.9) using (2.5), or immediately from (2.25),

$$E\tilde{\mathbf{W}} = E\mathbf{W}_{M/G/1} + \frac{1}{\gamma} + \frac{1}{1 - \lambda\beta_1} \int_{t=0}^{\infty} te^{-\gamma t} dR(t), \tag{3.10}$$

and

$$Var(\tilde{\mathbf{W}}) = Var(\mathbf{W}_{M/G/1}) + \frac{1}{\gamma^2} + \frac{1}{1 - \lambda\beta_1} \int_{t=0}^{\infty} t^2 e^{-\gamma t} dR(t) - \left(\frac{1}{1 - \lambda\beta_1} \int_{t=0}^{\infty} te^{-\gamma t} dR(t) \right)^2.$$

From (3.5) and (3.10), $E\mathbf{W}_{M/G/1} \leq E\mathbf{W} \leq E\tilde{\mathbf{W}}$, as expected.

An alternative derivation of (3.10) proceeds as follows. The waiting time of an individual customer is the sum of three terms, (different from the terms occurring in (3.9)), viz.:

- (i). the time from its arrival until its collection, $\tilde{\mathbf{W}}^{(i)}$;

- (ii). the waiting time of the supercustomer to which it belongs, $\tilde{\mathbf{W}}^{(ii)}$;
- (iii). the time from its admission until its service, $\tilde{\mathbf{W}}^{(iii)}$.

These three terms are dependent, however

$$E\tilde{\mathbf{W}} = E\tilde{\mathbf{W}}^{(i)} + E\tilde{\mathbf{W}}^{(ii)} + E\tilde{\mathbf{W}}^{(iii)}. \quad (3.11)$$

$\tilde{\mathbf{W}}^{(i)}$ is the length of a residual, negative exponentially distributed, collect interval. So

$$E\tilde{\mathbf{W}}^{(i)} = \frac{1}{\gamma}. \quad (3.12)$$

$E\tilde{\mathbf{W}}^{(ii)}$ does not equal $E\mathbf{W}$ because the supercustomer containing a tagged individual customer is not typical but is likely to have a long interarrival time and hence a short waiting time. However, applying Little's formula,

$$E\tilde{\mathbf{W}}^{(ii)} = \frac{E\tilde{\mathbf{N}}^{(ii)}}{\lambda}, \quad (3.13)$$

where $\tilde{\mathbf{N}}^{(ii)}$ denotes the number of individual customers belonging to waiting supercustomers. Furthermore

$$E\tilde{\mathbf{N}}^{(ii)} = \frac{E\mathbf{V}^{(ii)}}{\beta_1}, \quad (3.14)$$

where $\mathbf{V}^{(ii)}$ denotes the amount of work associated with individual customers belonging to waiting supercustomers. From Wolff [1989], p. 279 we have:

$$E\mathbf{V}^{(ii)} = \gamma E[\mathbf{W}\tau]. \quad (3.15)$$

Combining (2.14), (2.15), (3.8), (3.13), (3.14) and (3.15),

$$E\tilde{\mathbf{W}}^{(ii)} = E\mathbf{W}_{M/G/1} - \frac{\lambda\beta_1}{\gamma} + \frac{1}{1-\lambda\beta_1} \int_{t=0}^{\infty} te^{-\gamma t} dR(t). \quad (3.16)$$

$\tilde{\mathbf{W}}^{(iii)}$ is the amount of work that arrived during a past negative exponentially distributed collect interval. So

$$E\tilde{\mathbf{W}}^{(iii)} = \frac{\lambda\beta_1}{\gamma}. \quad (3.17)$$

Substituting (3.12), (3.16) and (3.17) in (3.11) yields (3.10).

4 THE NUMBER OF CUSTOMERS

In this section we focus our attention on the number of super- and individual customers in the system. Using the distributional form of Little's law, cf. Keilson & Servi [1990], it is easily seen that the generating function of \mathbf{N} , the number of *supercustomers* at an arbitrary time, and also at arrival epochs (PASTA), is given by $E\{z^{\mathbf{N}}\} = E\{e^{-\gamma(1-z)\mathbf{R}}\}$.

The remainder of this section is devoted to the number of *individual* customers in the system. Define \mathbf{A}_n and \mathbf{Z}_n as the number of individual customers which respectively have arrived but not yet been collected and have been collected but not yet departed, immediately after the departure of the n -th individual customer. Notice that $\{\mathbf{Z}_n, \mathbf{A}_n\}_{n \geq 1}$ is a two-dimensional Markov chain, whereas $\{\mathbf{Z}_n\}_{n \geq 1}$ and $\{\mathbf{A}_n\}_{n \geq 1}$ are not Markov chains. Let $\{\mathbf{Z}, \mathbf{A}\}$ be a vector with distribution the steady state distribution of this Markov chain. By letting $n \rightarrow \infty$ in the generating functions of $\{\mathbf{Z}_n, \mathbf{A}_n\}$ we will derive the generating function of $\{\mathbf{Z}, \mathbf{A}\}$.

For our analysis we need to define:

$\hat{\tau}_n :=$ Service time of individual customer n .

$\hat{\sigma}_n :=$ Interarrival time of the first supercustomer after and counted from the start of service of individual customer n .

If $\hat{\tau}_n < \hat{\sigma}_n$, then no supercustomers arrive during this service, otherwise at least one supercustomer arrives.

$\mu_n :=$ Number of individual customers which arrive during the service of individual customer n and are collected.

$\nu_n :=$ Number of individual customers which arrive after the last supercustomer arrival during the service of individual customer n . If no supercustomer arrival occurs, then ν_n is the total number of individual customers which arrive during the service of individual customer n .

$\zeta_n :=$ Number of individual customers which arrive between the end of service of individual customer $n - 1$ and the start of service of individual customer n . If $\mathbf{Z}_{n-1} > 0$, then $\zeta_n = 0$.

To illustrate these definitions, an example of the arrival and departure processes of customers is presented in figure 1.

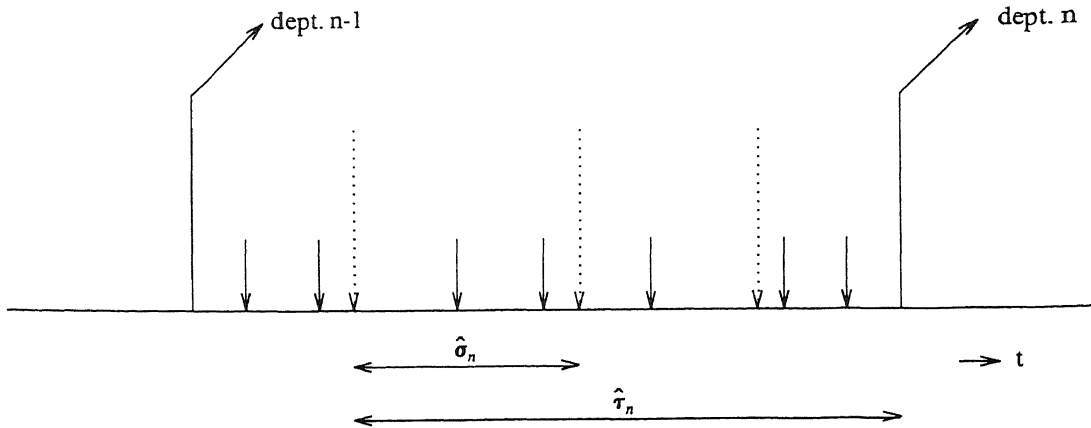


Figure 1: Arrivals and departures of individual customers, and arrivals of supercustomers (...). In this example, $\zeta_n = 2$, $\mu_n = 3$, $\nu_n = 2$.

With this notation, it is readily seen that the recurrence relations of $\{\mathbf{Z}_n, \mathbf{A}_n\}_{n \geq 1}$ are given by

$$\{\mathbf{Z}_{n+1}, \mathbf{A}_{n+1}\} = \begin{cases} \{\mathbf{Z}_n - 1, \mathbf{A}_n + \nu_{n+1}\} & \text{if } \mathbf{Z}_n \geq 1, \hat{\tau}_{n+1} < \hat{\sigma}_{n+1} \\ \{\mathbf{Z}_n - 1 + \mathbf{A}_n + \mu_{n+1}, \nu_{n+1}\} & \text{if } \mathbf{Z}_n \geq 1, \hat{\tau}_{n+1} \geq \hat{\sigma}_{n+1} \\ \{\mathbf{A}_n + \zeta_{n+1} - 1 + \mu_{n+1}, \nu_{n+1}\} & \text{if } \mathbf{Z}_n = 0 \end{cases}$$

Now define for $|z| \leq 1, |q| \leq 1, n = 1, 2, \dots$, the generating functions $\Phi_n(z, q) := E\{z^{\mathbf{Z}_n} q^{\mathbf{A}_n}\}$. Using that $\mathbf{Z}_n, \mathbf{A}_n, \zeta_{n+1}, I_{\{\mathbf{Z}_n \geq 1\}}$ and $I_{\{\mathbf{Z}_n = 0\}}$ are independent of $\nu_{n+1}, \mu_{n+1}, I_{\{\hat{\tau}_{n+1} \geq \hat{\sigma}_{n+1}\}}$ and $I_{\{\hat{\tau}_{n+1} < \hat{\sigma}_{n+1}\}}$, it follows for $|z| \leq 1, |q| \leq 1$ that

$$\begin{aligned} \Phi_{n+1}(z, q) &= \frac{1}{z} E\{z^{\mathbf{Z}_n} q^{\mathbf{A}_n} I_{\{\mathbf{Z}_n \geq 1\}}\} E\{q^{\nu_{n+1}} I_{\{\hat{\tau}_{n+1} < \hat{\sigma}_{n+1}\}}\} + \quad (4.1) \\ &\quad \frac{1}{z} E\{z^{\mathbf{Z}_n + \mathbf{A}_n} I_{\{\mathbf{Z}_n \geq 1\}}\} E\{z^{\mu_{n+1}} q^{\nu_{n+1}} I_{\{\hat{\tau}_{n+1} \geq \hat{\sigma}_{n+1}\}}\} + \\ &\quad \frac{1}{z} E\{z^{\mathbf{A}_n + \zeta_{n+1}} I_{\{\mathbf{Z}_n = 0\}}\} E\{z^{\mu_{n+1}} q^{\nu_{n+1}}\}. \end{aligned}$$

We consider each of the three terms in the right-hand side of (4.1) in turn. The calculations involved can be found in detail in Borst et al. [1991].

$$E\{q^{\nu_{n+1}} I_{\{\hat{\tau}_{n+1} < \hat{\sigma}_{n+1}\}}\} = \beta(\gamma + \lambda(1 - q)), \quad (4.2)$$

$$E\{z^{\mu_{n+1}} q^{\nu_{n+1}} I_{\{\hat{\tau}_{n+1} \geq \hat{\sigma}_{n+1}\}}\} = \frac{\gamma[\beta(\lambda(1 - z)) - \beta(\gamma + \lambda(1 - q))]}{\gamma + \lambda(z - q)}, \quad (4.3)$$

$$E\{z^{\mu_{n+1}} q^{\nu_{n+1}}\} = \frac{\gamma\beta(\lambda(1 - z))}{\gamma + \lambda(z - q)} + \frac{\lambda(z - q)\beta(\gamma + \lambda(1 - q))}{\gamma + \lambda(z - q)}, \quad (4.4)$$

$$E\{z^{\mathbf{A}_n + \zeta_{n+1}} I_{\{\mathbf{Z}_n = 0\}}\} = \frac{\gamma}{\gamma + \lambda(1 - z)} E\{z^{\mathbf{A}_n} I_{\{\mathbf{Z}_n = 0\}} I_{\{\mathbf{A}_n > 0\}}\} + \frac{\gamma z}{\gamma + \lambda(1 - z)} E\{I_{\{\mathbf{Z}_n = 0\}} I_{\{\mathbf{A}_n = 0\}}\}. \quad (4.5)$$

We next let $n \rightarrow \infty$; note that $\{\mathbf{Z}_n, \mathbf{A}_n\}_{n \geq 1}$ is an irreducible, aperiodic Markov chain, so that $\Phi(z, q) := \lim_{n \rightarrow \infty} \Phi_n(z, q)$ and related limits exist. Substituting (4.2) - (4.5) into (4.1) we obtain the functional equation

$$\begin{aligned} & \Phi(z, q)[z - \beta(\gamma + \lambda(1 - q))] = \\ & - \Phi(0, q)\beta(\gamma + \lambda(1 - q)) \quad (4.6) \\ & + \Phi(z, z) \frac{\gamma}{\gamma + \lambda(z - q)} [\beta(\lambda(1 - z)) - \beta(\gamma + \lambda(1 - q))] \\ & - \Phi(0, 0) \frac{\gamma}{\gamma + \lambda(z - q)} \frac{[1 - z][\gamma\beta(\lambda(1 - z)) + \lambda(z - q)\beta(\gamma + \lambda(1 - q))]}{\gamma + \lambda(1 - z)} \\ & + \Phi(0, z) \frac{\gamma}{\gamma + \lambda(z - q)} \frac{(\gamma + \lambda(1 - q))\beta(\gamma + \lambda(1 - q)) - \lambda(1 - z)\beta(\lambda(1 - z))}{\gamma + \lambda(1 - z)}. \end{aligned}$$

With up/down crossing arguments and the PASTA property we see that $\Phi(q, q)$ is equal to the generating function of the total number of individual customers at an arbitrary time, a function which we obtained earlier, cf. (2.20) - (2.23). Taking $z = q$ in (4.6), we thus find

$$\Phi(0, q) = \frac{\gamma}{\lambda(\gamma + \lambda)} [(\gamma + \lambda)r(\gamma + \lambda(1 - q)) - \gamma r(\gamma + \lambda)]. \quad (4.7)$$

Substituting the expressions for $\Phi(z, z)$, $\Phi(0, 0)$ and $\Phi(0, q)$ in (4.6), together with $r(\cdot)$ as given by (2.10), we obtain a closed form expression for $\Phi(z, q)$.

5 THE BUSY PERIOD

For the sojourn time of a supercustomer we were able to derive the LST, starting from recurrence relation (2.1). This relation is a typical starting point in $M/G/1$ analysis concerning waiting and sojourn processes. A similar starting point in the ordinary $M/G/1$ queue for busy period analysis is the branching argument, cf. Cohen [1982] p. 249. Unfortunately, this argument does not apply to our model. This is due to:

- The distribution of the service time of a supercustomer initiating a busy period is hard to determine because the interarrival time of this customer is atypical. The fact that the previous busy period has ended during its interarrival interval suggests that this interarrival interval is relatively large.
- The length of a (sub)busy period initiated by a supercustomer arriving during the service of the first supercustomer in a busy period depends on the number of supercustomers arriving during that service.

While deriving the LST of the busy period seems a difficult problem, the average length of a busy period is easily obtained using a balancing argument. With EB the mean busy period length, $E\mathbf{I} = \frac{1}{\gamma}$ the mean idle period length and using

$$\frac{EB}{EB + E\mathbf{I}} = \lambda\beta_1 = \rho,$$

the mean busy period length is given by

$$EB = \frac{\lambda\beta_1/\gamma}{1 - \rho},$$

which is the same as EB_I , the mean busy period length in the corresponding $M/G/1$ queue without dependence. Note that a busy period can have length zero.

Of more interest are ‘real’ busy periods, viz. busy periods initiated by a supercustomer with positive service time. This conditioned mean busy period length is obviously larger than EB and EB_I .

Define:

- τ^1 := service time of a customer initiating a new busy period.
- $\tilde{\mathbf{B}}$:= the length of a non-zero busy period.

Then,

$$E\mathbf{B} = E\tilde{\mathbf{B}} \cdot Pr\{\tau^1 > 0\}.$$

Using

$$\begin{aligned} Pr\{\tau^1 > 0\} &= Pr\{\tau > 0 \mid \mathbf{W} = 0\} = 1 - Pr\{\tau = 0 \mid \mathbf{W} = 0\} \\ &= 1 - \frac{Pr\{\tau = 0, \mathbf{W} = 0\}}{Pr\{\mathbf{W} = 0\}} = 1 - \frac{Pr\{\mathbf{R} = 0\}}{1 - \rho} \end{aligned}$$

and

$$Pr\{\mathbf{R} = 0\} = \frac{\gamma}{\gamma + \lambda} r(\gamma + \lambda) \quad (\text{see Remark 2.1}),$$

we obtain

$$E\tilde{\mathbf{B}} = \frac{\lambda\beta_1}{\gamma} \frac{1}{1 - \rho - \frac{\gamma}{\gamma + \lambda} r(\gamma + \lambda)}. \quad (5.1)$$

Remark 5.1 Letting $\gamma \rightarrow \infty$ and using Remark 2.2 gives as expected $E\tilde{\mathbf{B}} \rightarrow \frac{\beta_1}{1 - \rho}$, the mean of $\mathbf{B}_{M/G/1}$, the busy period length in an ordinary M/G/1 queue. □

Remark 5.2 Letting $\omega = \gamma$ in (2.3) and using the convexity of $r(\cdot)$ gives $E\tilde{\mathbf{B}} > E\mathbf{B}_{M/G/1}$. □

Remark 5.3 Although $E\tilde{\mathbf{B}}$ is larger than $E\mathbf{B}_{M/G/1}$ and $E\mathbf{B}_I$, we expect $cv(\tilde{\mathbf{B}})$, the coefficient of variation of the length of a non-zero busy period, to be smaller than $cv(\mathbf{B}_{M/G/1})$ and $cv(\mathbf{B}_I)$. Due to the positive correlation between service and interarrival time of a supercustomer the injection of workload is more regulated than in an ordinary M/G/1 queue. This regulation has a stabilizing effect on the busy period length. Moreover, the stronger the correlation, the smaller we expect $cv(\tilde{\mathbf{B}})$ to be. Our conjectures, also stated by Hadidi [1981] for a similar model, are supported by simulation results which we present in the next section. □

6 NUMERICAL RESULTS

In this section we present some numerical results to see the quantitative effects of arrival and traffic intensities in our model. We consider the influence of different service time distributions and we also consider which effect higher moments of the service time distribution have. Finally, we support our claims about the coefficient of variation of the busy period length.

The results have mostly been obtained by numerical evaluation of the infinite product (2.9), its derivative (2.11) and well known formulas for the ordinary $M/G/1$ queue. The infinite product (2.9) and the infinite sum (2.11) converge very fast unless ρ is close to one; it is easily verified that the difference between the k -th term in (2.9) and 1 is of order $O(\rho^k)$ and the k -th term in (2.11) is of order $O(\rho^k)$ for $k = 1, 2, \dots$. We have only taken recourse to simulation for determining the coefficient of variation of the busy period length.

Waiting times of supercustomers and individual customers.

In tables 1a, 1b and 1c we compare the mean waiting times of supercustomers, EW , of customers in the corresponding $M/G/1$ queue without dependence, EW_I , and of individual customers, $E\tilde{W}$. We also compare these mean waiting times with the mean waiting times of customers in the corresponding $M/G/1$ queue without collection, $EW_{M/G/1}$ (in the tables represented by $\gamma = \infty$). Here and in the rest of this section λ is fixed with value 1.

In table 1c the service time of a message is with probability $\frac{1}{4}$ exponentially distributed with parameter μ_1 and with probability $\frac{3}{4}$ exponentially distributed with parameter μ_2 . $\mu_1 = 1$ and $\mu_2 = 3$ for $\rho = 0.5$; $\mu_1 = \frac{5}{9}$ and $\mu_2 = \frac{5}{3}$ for $\rho = 0.9$.

Comparing EW and EW_I in 1a-c we conclude that the positive correlation between interarrival and service times leads to a reduction of mean waiting times. The reduction is particularly strong in heavy traffic. A similar observation has been made by Hadidi [1981] for the dependence structure displayed in formula (1.5). Tables 1a, 1b and 1c also show that the influence of the service time distribution of individual customers decreases when γ approaches 0. In fact, it can be seen from (1.2) that for any service time distribution $B(\cdot)$, $\lim_{\gamma \downarrow 0} E\{e^{-\gamma\omega\mathcal{T}_i}\} = (1 + \lambda\beta_1\omega)^{-1}$, i.e. the distribution of the scaled service time of a supercustomer converges to the negative exponential distribution with mean $\lambda\beta_1$.

Table 1

Comparison of the mean waiting times with those in the corresponding $M/G/1$ queue without dependence.

TABLE 1a Exponential service								
	$\rho = 0.5$				$\rho = 0.9$			
γ	$cor.(\sigma, \tau)$	EW	EW_I	$E\tilde{W}$	$cor.(\sigma, \tau)$	EW	EW_I	$E\tilde{W}$
0.1	0.9129	3.392	5.500	16.292	0.9129	23.197	89.100	34.875
0.5	0.7071	0.987	1.500	3.474	0.7071	9.951	24.300	12.157
1	0.5774	0.701	1.000	1.902	0.5774	8.729	16.200	9.798
2	0.4472	0.574	0.750	1.148	0.4472	8.290	12.150	8.811
4	0.3333	0.524	0.625	0.797	0.3333	8.152	10.125	8.408
10	0.2182	0.504	0.550	0.609	0.2182	8.109	8.910	8.210
∞	0	0.500	0.500	0.500	0	8.100	8.100	8.100

TABLE 1b Deterministic service								
	$\rho = 0.5$				$\rho = 0.9$			
γ	$cor.(\sigma, \tau)$	EW	EW_I	$E\tilde{W}$	$cor.(\sigma, \tau)$	EW	EW_I	$E\tilde{W}$
0.1	0.9535	3.213	5.250	16.176	0.9535	20.494	85.050	32.321
0.5	0.8165	0.790	1.250	3.330	0.8165	6.452	20.250	8.718
1	0.7071	0.491	0.750	1.731	0.7071	4.937	12.150	6.036
2	0.5774	0.346	0.500	0.943	0.5774	4.327	8.100	4.859
4	0.4472	0.282	0.375	0.565	0.4472	4.121	6.075	4.378
10	0.3015	0.255	0.300	0.361	0.3015	4.060	4.860	4.161
∞	0	0.250	0.250	0.250	0	4.050	4.050	4.050

TABLE 1c Hyper-2 exponential service								
	$\rho = 0.5$				$\rho = 0.9$			
γ	$cor.(\sigma, \tau)$	EW	EW_I	$E\tilde{W}$	$cor.(\sigma, \tau)$	EW	EW_I	$E\tilde{W}$
0.1	0.9393	3.521	5.666	16.374	0.9393	25.160	91.800	36.756
0.5	0.7746	1.128	1.666	3.589	0.7746	12.463	27.000	14.647
1	0.6547	0.853	1.166	2.040	0.6547	11.364	18.900	12.426
2	0.5222	0.734	0.916	1.302	0.5222	10.974	14.850	11.493
4	0.3974	0.689	0.791	0.961	0.3974	10.850	12.825	11.105
10	0.2641	0.671	0.671	0.775	0.2641	10.809	11.610	10.910
∞	0	0.666	0.666	0.666	0	10.800	10.800	10.800

Finally we see that $E\bar{W}$ converges slower towards $EW_{M/G/1}$ than EW for $\gamma \rightarrow \infty$, but most of the difference is due to the remaining collect interval.

The influence of higher moments

In an ordinary $M/G/1$ queue, the influence of the service time distribution on mean waiting time is limited to its first and second moment. The question arises whether that is the case in our model. Formula (2.14) shows that $r'(\gamma)$ contributes to the mean waiting time and therefore we would suspect that the whole service time distribution plays a role. Tables 2a and 2b indicate that this conjecture is correct but also that the influence of higher moments is almost negligible. In these tables we consider mixtures of exponential distributions for the service time of individual customers.

In table 2 $EW1$ and $EW2$ are the mean waiting times for a supercustomer composed of individual customers with service time distribution mixtures 1 and 2, respectively.

Table 2

The influence of higher service time moments on the mean waiting time.

γ	$EW1$	$EW2$
0.1	5.5479	5.5960
0.5	3.4816	3.5304
1	3.3154	3.3451
2	3.2559	3.2697
4	3.2344	3.2393
10	3.2252	3.2260

	β_1	β_2	β_3
Mix.1	0.5	3.222	38.407
Mix.2	0.5	3.222	52.345

γ	$EW1$	$EW2$
0.1	71.597	72.293
0.5	64.657	64.943
1	64.291	64.426
2	64.173	64.226
4	64.113	64.149
10	64.116	64.119

	β_1	β_2	β_3
Mix.1	0.9	12.822	307.207
Mix.2	0.9	12.822	423.554

Table 3

The busy period length, mean and coefficient of variation.

TABLE 3a Exponential service								
	$\rho = 0.5$				$\rho = 0.9$			
γ	$E\tilde{\mathbf{B}}$	$cv(\tilde{\mathbf{B}})$	$E\mathbf{B}_I$	$cv(\mathbf{B}_I)$	$E\tilde{\mathbf{B}}$	$cv(\tilde{\mathbf{B}})$	$E\mathbf{B}_I$	$cv(\mathbf{B}_I)$
0.1	10.076	0.954	10.000	1.844	90.057	1.473	90.000	4.583
0.5	2.439	1.230	2.000	2.236	20.423	2.763	18.000	5.385
1	1.603	1.407	1.000	2.646	13.356	3.546	9.000	6.245
2	1.233	1.557	0.500	3.317	10.517	3.991	4.500	7.681
4	1.080	1.652	0.250	4.359	9.465	4.123	2.250	9.950
10	1.016	1.711	0.100	6.557	9.089	4.266	0.900	14.799
∞	1.000	1.732	0.000	∞	9.000	4.359	0.000	∞

TABLE 3b Deterministic service								
	$\rho = 0.5$				$\rho = 0.9$			
γ	$E\tilde{\mathbf{B}}$	$cv(\tilde{\mathbf{B}})$	$E\mathbf{B}_I$	$cv(\mathbf{B}_I)$	$E\tilde{\mathbf{B}}$	$cv(\tilde{\mathbf{B}})$	$E\mathbf{B}_I$	$cv(\mathbf{B}_I)$
0.1	10.053	0.894	10.000	1.789	90.012	1.168	90.000	4.472
0.5	2.374	0.983	2.000	2.000	19.525	2.043	18.000	4.899
1	1.534	0.995	1.000	2.236	12.324	2.540	9.000	5.385
2	1.167	0.994	0.500	2.646	9.702	2.856	4.500	6.245
4	1.032	1.002	0.250	3.317	9.061	3.002	2.250	7.811
10	1.001	0.999	0.100	4.796	9.000	2.971	0.900	10.909
∞	1.000	1.000	0.000	∞	9.000	3.000	0.000	∞

TABLE 3c Hyper-2 Exponential service (same as in TABLE 1c)								
	$\rho = 0.5$				$\rho = 0.9$			
γ	$E\tilde{\mathbf{B}}$	$cv(\tilde{\mathbf{B}})$	$E\mathbf{B}_I$	$cv(\mathbf{B}_I)$	$E\tilde{\mathbf{B}}$	$cv(\tilde{\mathbf{B}})$	$E\mathbf{B}_I$	$cv(\mathbf{B}_I)$
0.1	10.086	1.010	10.000	1.880	90.090	1.591	90.000	4.655
0.5	2.463	1.390	2.000	2.380	20.720	3.284	18.000	5.686
1	1.625	1.645	1.000	2.887	13.614	4.068	9.000	6.758
2	1.249	1.863	0.500	3.697	10.675	4.492	4.500	8.505
4	1.089	1.973	0.250	4.933	9.548	4.831	2.250	11.210
10	1.019	2.052	0.100	7.506	9.107	4.852	0.900	16.902
∞	1.000	2.082	0.000	∞	9.000	5.066	0.000	∞

The busy period

In Section 5 we suggested that in our model the coefficient of variation for real busy periods, $cv(\tilde{\mathbf{B}})$, would be smaller than $cv(\mathbf{B}_{M/G/1})$ and $cv(\mathbf{B}_I)$. As explained there, analytical and numerical results are not available, so to obtain column $cv(\tilde{\mathbf{B}})$ in tables 3a, 3b and 3c below we have used a simulation. The simulation was performed with the queueing simulation software package Q+, running the process for 10^6 time units. $E\tilde{\mathbf{B}}$ has been obtained using formula (5.1). In tables 3a, 3b, 3c $cv(\tilde{\mathbf{B}})$ is smaller than $cv(\mathbf{B}_{M/G/1})$ and $cv(\mathbf{B}_I)$, supporting the conjectures made in Remark 5.3.

7 CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

A detailed exact analysis has been presented of a variant of an $M/G/1$ queue in which a special form of positive correlation exists between interarrival and service times. Explicit expressions are derived for (transforms of) sojourn time, waiting time and queue length distributions. The mean waiting time is proven to be smaller than the mean waiting time in the uncorrelated situation. Numerical results show that a strong positive correlation between interarrival and service times may lead to a very strong reduction of mean waiting times, in particular in heavy traffic. The latter observation agrees with results obtained in Boxma [1979] for a somewhat similar model. That paper studies two queues in series with *identical*, generally distributed, service times of any customer at the two queues. Obviously interarrival and service times at the second queue are now also identical when the first queue is not empty. Numerical and asymptotic results in Boxma [1979] expose the significant reduction in mean and variance of the sojourn time at the second queue, that occurs in heavy traffic. A very recent report of Cidon et al. [1991b] also analyses a single server queue with service times depending on the preceding interarrival time, these two quantities obeying some form of linear proportionality. For that case too, significant reductions of waiting times can occur. The models of Cidon et al. [1991b], of Conolly & Choo [1979], and of the present paper can be integrated to allow a more general and flexible dependence modelling, which is still amenable to an exact analysis. Indeed, in Boxma & Combé [1993] the following generalization is analysed: work arrives according to a process with stationary non-negative independent increments and is collected and delivered to the server at exponential intervals.

Acknowledgement The authors are indebted to Professor J.W. Cohen for several interesting discussions; further they are grateful to the referees for several useful suggestions and in particular for mentioning the papers of Coleman [1973] and Takahashi [1971].

REFERENCES

- [1992] Abate, J., Whitt, W. (1992). The Fourier-Series Method for Inverting Transforms of Probability Distributions. *Queueing Systems* **10**, 5-87.
- [1984] Ali, O.M.E., Neuts, M.F. (1984). A Service System with Two Stages of Waiting and Feedback of Customers. *J. Appl. Prob.* **21**, 404-413.
- [1991] Borst, S.C., Boxma, O.J., Combé, M.B. (1991). An $M/G/1$ Queue with Dependence between Interarrival and Service Times. CWI Report BS-R9125.
- [1992] Borst, S.C., Boxma, O.J., Combé, M.B. (1992). Collection of Customers: a Correlated $M/G/1$ Queue. *Perf. Eval. Review* **20**, 47-59.
- [1992] Borst, S.C., Combé, M.B. (1992). Busy Period Analysis of a Correlated Queue. *J. Appl. Prob.* **29**, 482-483.
- [1979] Boxma, O.J. (1979). On a Tandem Queue with Identical Service Times at Both Counters, II. *Adv. Appl. Prob.* **11**, 644-659.
- [1989] Boxma, O.J. (1989). Workloads and Waiting Times in Single-Server Systems with Multiple Customer Classes. *Queueing Systems* **5**, 185-214.
- [1993] Boxma, O.J., Combé, M.B. (1993). The Correlated $M/G/1$ Queue. CWI Report BS-R9303.
- [1992] Boxma, O.J., Levy, H., Yechiali, U. (1992). Cyclic Reservation Schemes for Efficient Operation of Multiple-Queue Single-Server Systems. *Annals of Operations Research* **35**, 187-208, *Special Issue on Stochastic Modelling of Telecommunication Systems*.

- [1991a] Cidon, I., Guérin, R., Khamisy, A., Sidi, M. (1991). On Queues with Inter-Arrival Times Proportional to Service Times. Report Technion, EE PUB No. 811, December 1991.
- [1991b] Cidon, I., Guérin, R., Khamisy, A., Sidi, M. (1991). Analysis of a Correlated Queue in Communication Systems. Report Technion, EE PUB No. 812, December 1991.
- [1982] Cohen, J.W. (1982). *The Single Server Queue*. (North-Holland, Amsterdam, 2nd ed.).
- [1973] Coleman, R.D. (1973). Use of a Gate to Reduce the Variance of Delays in Queues with Random Service. *Bell Syst. Techn. J.* **52**, 1403-1422.
- [1968] Conolly, B.W. (1968). The Waiting Time for a Certain Correlated Queue. *Oper. Res.* **15**, 1006-1015.
- [1979] Conolly, B.W., Choo, Q.H. (1979). The Waiting Time Process for a Generalized Correlated Queue with Exponential Demand and Service. *SIAM J. Appl. Math.* **37**, 263-275.
- [1969] Conolly, B.W., Hadidi, N. (1969). A Correlated Queue. *J. Appl. Prob.* **6**, 122-136.
- [1985] Fuhrmann, S.W., Cooper, R.B. (1985). Stochastic Decompositions in the $M/G/1$ Queue with Generalized Vacations. *Oper. Res.* **33**, 1117-1129.
- [1981] Hadidi, N. (1981). Queues with Partial Correlation. *SIAM J. Appl. Math.* **40**, 467-475.
- [1985] Hadidi, N. (1985). Further Results on Queues with Partial Correlation. *Oper. Res.* **33**, 203-209.
- [1990] Keilson, J., Servi, L.D. (1990). The Distributional Form of Little's Law and the Fuhrmann-Cooper Decomposition. *Oper. Res. Letters* **9**, 239-247.
- [1987] Langaris, C. (1987). Busy-Period Analysis of a Correlated Queue with Exponential Demand and Service. *J. Appl. Prob.* **24**, 476-485.

- [1962] Loynes, R.M. (1962). The Stability of a Queue with Non-Independent Interarrival and Service Times. *Proc. Cambridge Phil. Soc.* **58**, 497-520.
- [1989] Nassehi, M.M. (1989). CRMA: An Access Scheme for High-Speed LANs and MANs. Report IBM Zürich.
- [1971] Takahashi, Y. (1971). Queueing Systems with Gates. *J. Oper. Res. Soc. Jap.* **14**, 103-126.
- [1989] Wolff, R.W. (1989). *Stochastic Modeling and the Theory of Queues*. (Prentice Hall, Englewood Cliffs, NJ).

Received: 10/21/1991
Revised: 5/27/1992
Accepted: 3/3/1993

Recommended by Yukio Takahashi, Editor