An M/G/1 queue with multiple types of feedback and gated vacations

Onno J. Boxma[1], Uri Yechiali

ABSTRACT This paper considers a single-server queue with Poisson arrivals and multiple customer feedbacks. If the first service attempt of a newly arriving customer is not successful, he returns to the end of the queue for another service attempt, with a different service time distribution. He keeps trying in this manner (as an 'old' customer) until his service is successful. The server operates according to the 'gated vacation' strategy: when it returns from a vacation to find K (new and old) customers, it renders a single service attempt to each of them and takes another vacation, etc. We study the joint queue length process of new and old customers, as well as the waiting time distribution of customers. Some extensions are also discussed.

1 Introduction

This paper is dedicated to Professor Joe Gani. The applied probability community is forever indebted to him, for his vigorous and unselfish promotion of applied probability as a scientific discipline.
Consider a dedicated editor-in-chief who devotes sessions to handling tasks for his scientific journal. He distinguishes between new and old tasks. New tasks are papers/reports that he has received from his managing editor since the start of the previous working session. He decides on acceptance, or on the choice of a referee, etc. Some of the new tasks cannot yet be handled, as the editor-in-chief needs to "sleep on it" or to get additional information from his managing editor. Such deferred tasks will be treated in the next session as old tasks. At the end of the session the editor sends a message to the managing editor and takes a vacation until a reply comes. This situation can be modeled by a feedback queue with so-called gated vacations - which is the subject of this paper.
Other situations modelled by such a feedback queue are that of a machine processing parts, X-ray taking, and the like. After a session, the processed

[1]This paper is based on an invited presentation at the Applied Probability conference in honour of J. Gani and T.E. Hannan, Athens, March 1995.

parts (or X-rays) are inspected and it is decided which have to be redone. In the model studied in this paper we allow different service time distributions for new and old tasks, and for successful and unsuccessful tasks. This enables us, for example, to model a situation where a job can be processed either in one session, or be divided over two (or more) sessions. Our model contains the ordinary M/G/1 queue with gated vacations as a special case (cf. Takagi [8]); it also encompasses the M/G/1 queue with Bernoulli feedback (cf. Takács [7]), and its generalization to an M/G/1 queue with Bernoulli feedback and gated vacations (all service time distributions being the same), cf. Takine et al. [9].

The paper is organized as follows. The model is described in detail below. In Section 2 we determine the joint distribution of the number of old and new customers at session beginnings. The steady-state joint queue length distribution is studied in Section 3, and the waiting time distribution in Section 4.

The model

Customers arrive at a single-server $M/G/1$-type queue according to a Poisson process with rate λ. A newly arriving customer receives a 'successful' service with probability p, or an 'unsuccessful' service with probability $1-p$. The duration of a successful service attempt is S (with mean s, pdf $S(\cdot)$ and LST $\tilde{S}(\cdot)$), while the duration of an unsuccessful service attempt is U (mean u, pdf $U(\cdot)$ and LST $\tilde{U}(\cdot)$). If a service attempt is successful the customer leaves the system, whereas if the attempt is unsuccessful, the customer is immediately fed back to the end of the queue. ¿From that moment on, such a customer is an 'old' one. If a customer has been fed back *at least once* then, on each of his following service attempts, with probability p_1 he receives a successful service with duration S_1 (mean s_1, pdf $S_1(\cdot)$, LST $\tilde{S}_1(\cdot)$), and with probability $1-p_1$ he receives an unsuccessful service U_1 (mean u_1, pdf $U_1(\cdot)$ and LST $\tilde{U}_1(\cdot)$).

A generalization of the above model to M types of service re-attempts is straightforward to formulate, and its analysis is not fundamentally more complicated; it will not be discussed in this paper.

The server operates according to the gated vacation strategy. When the server returns from a vacation to find, say, N new customers and O old ones, it starts a "session" and renders each one of those customers a single service attempt (which may be successful or not) and then leaves for another vacation whose duration V has mean v, pdf $V(\cdot)$ and LST $\tilde{V}(\cdot)$. If the queue is empty upon returning from a vacation, the server immediately leaves for another vacation, etc. (Note that we have assumed that a customer immediately leaves after a successful service. In some applications it may be more realistic that a successful customer is only released after an inspection at the beginning of the server vacation. The analysis to be presented below can easily handle that.)

We impose all the usual independence assumptions between the arrival process, successful and unsuccessful services, and vacations. Given that a customer's first service attempt is unsuccessful (with duration U), let Y be the number of his additional unsuccessful attempts *before* he leaves the system. Clearly, $P(Y = n) = (1 - p_1)^n p_1$, $n = 0, 1, 2, \cdots$. Thus, the expected total amount of service provided to a customer by the time he successfully leaves the system is

$$ps + (1 - p)\{u + E[\sum_{j=1}^{Y} U_{1j}] + s_1\}$$

where U_{1j} are independent, identically distributed as U_1.
Since $EY = \frac{1-p_1}{p_1}$, by Wald's theorem, $E\left[\sum_{j=1}^{Y} U_{1j}\right] = \left(\frac{1}{p_1} - 1\right) u_1$. Thus, the system load is

$$\rho := \lambda \left[ps + (1 - p)\left\{u + \left(\frac{1}{p_1} - 1\right) u_1 + s_1\right\}\right] . \tag{1.1}$$

An interpretation of ρ is as follows. Let $B_N = pS + (1 - p)U$ denote the duration of a service attempt of a *new* customer (mean b_N, pdf $B_N(\cdot)$, LST $\tilde{B}_N(\cdot)$), and let $B_O = p_1 S_1 + (1 - p_1)U_1$ denote the duration of a service attempt of an *old* customer (mean b_O, pdf $B_O(\cdot)$, LST $\tilde{B}_O(\cdot)$). Clearly,

$$b_N = ps + (1 - p)u, \quad b_O = p_1 s_1 + (1 - p_1)u_1 .$$

Each customer makes one service attempt (of mean duration b_N) as a *new* customer. With probability $(1 - p)$ it is unsuccessful, and then it obtains an average of $\frac{1}{p_1}$ additional service attempts as an *old* customer, each attempt with mean duration b_O. Thus, the mean load of each customer on the system is $b_N + \frac{1-p}{p_1} b_O$, and the overall system load ρ is as obtained in (1.1). We assume that $\rho < 1$.

Remark. In [3] polling models with customer routing are considered; this includes feedback queues, with different service time distributions at successive visits. That paper sketches a framework for analyzing joint queue length distributions in such cases. The approach is based on branching processes, which are basically also at the heart of the present paper.

2 The joint queue length distribution at session beginnings

Define
T_n := the time instant when the nth vacation ends.
N_n := number of 'new' customers at T_n. (Those customers have arrived

during the time interval $(T_{n-1}, T_n]$.)
$O_n :=$ number of 'old' customers at T_n.
The joint probability generating function (PGF) of N_n and O_n is

$$G_n(z,w) := E[z^{N_n} w^{O_n}], \qquad n = 1,2,\cdots,\ |z| \le 1,\ |w| \le 1\ . \tag{2.1}$$

We shall derive a recursive formula for $G_n(z,w)$ that will lead to the limiting distribution of N_n and O_n. Let $X(N_n)$ (respectively $X_1(O_n)$, $A(V_n)$) denote the number of new arrivals during the service time of the N_n 'new' customers (respectively the service time of the O_n 'old' customers, and the nth vacation V_n).
Let $F(N_n)$ (respectively $F_1(O_n)$) denote the number of customers fed back out of N_n (respectively, O_n). Then,

$$\begin{aligned} N_{n+1} &= X(N_n) + X_1(O_n) + A(V_n), \\ O_{n+1} &= F(N_n) + F_1(O_n). \end{aligned} \tag{2.2}$$

We claim:
Proposition 1. For $|z| \le 1, |w| \le 1$:

$$\begin{aligned} G_{n+1}(z,w) &= \tilde{V}\big(\lambda(1-z)\big) \cdot G_n\Big(p\tilde{S}\big[\lambda(1-z)\big] + (1-p)w\tilde{U}\big[\lambda(1-z)\big], \\ &\qquad p_1\tilde{S}_1\big[\lambda(1-z)\big] + (1-p_1)w\tilde{U}_1\big[\lambda(1-z)\big]\Big)\ . \end{aligned} \tag{2.3}$$

Proof
If $N_n = j$ and $O_n = k$, then with probability

$$\binom{j}{m} p^m (1-p)^{j-m} \binom{k}{m_1} p_1^{m_1} (1-p_1)^{k-m_1}\ ,$$

there will be m type S (m_1 type S_1) successful services out of $N_n = j$ (out of $O_n = k$) attempts, and $j-m$ type U ($k-m_1$ type U_1) unsuccessful services.
The total service duration of the $N_n = j$ and $O_n = k$ customers will have a convolution pdf $\big[S^{m*} * U^{(j-m)*} * S_1^{m_1 *} * U_1^{(k-m_1)*}\big](\cdot)$. If that convoluted service lasts t units of time, there will be ℓ new arrivals during that time with probability $e^{-\lambda t}\frac{(\lambda t)^\ell}{\ell!}$ $(\ell = 0,1,2,\cdots)$. Thus,

$$\begin{aligned} &E\big[z^{N_{n+1}} w^{O_{n+1}} \mid N_n = j, O_n = k\big] = E\big[z^{A(V_n)}\big] \cdot \\ &\cdot \sum_{m=0}^{j} \sum_{m_1=0}^{k} \binom{j}{m} p^m (1-p)^{j-m} w^{j-m} \binom{k}{m_1} p_1^{m_1} (1-p_1)^{k-m_1} w^{k-m_1} \cdot \\ &\cdot \sum_{\ell=0}^{\infty} z^\ell \int_{t=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^\ell}{\ell!} d\big[S^{m*} * U^{(j-m)*} * S_1^{m_1 *} * U_1^{(k-m_1)*}\big](t) \end{aligned}$$

$$= \tilde{V}\big[\lambda(1-z)\big]\Big[p\tilde{S}\big[\lambda(1-z)\big] + (1-p)w\tilde{U}\big[\lambda(1-z)\big]\Big]^j \cdot$$
$$\cdot \Big[p_1\tilde{S}_1\big[\lambda(1-z)\big] + (1-p_1)w\tilde{U}_1\big[\lambda(1-z)\big]\Big]^k \ . \tag{2.4}$$

Now, taking expectation w.r.t. N_n and O_n, we obtain equation (2.3), Q.E.D. Let N and O be the random variables with joint distribution the joint *limiting distribution* of N_n and O_n. Then, for $|z| \leq 1$, $|w| \leq 1$,

$$\begin{aligned} G(z,w) = E[z^N w^O] &= \tilde{V}\big[\lambda(1-z)\big] \cdot \\ &\cdot G\Big(p\tilde{S}\big[\lambda(1-z)\big] + (1-p)w\tilde{U}\big[\lambda(1-z)\big]\ , \\ &p_1\tilde{S}_1\big[\lambda(1-z)\big] + (1-p_1)w\tilde{U}_1\big[\lambda(1-z)\big]\Big) \ . \end{aligned} \tag{2.5}$$

Remark. A fast way to understand (2.5) is to use the concept of branching processes. According to this concept one can interpret the factor

$$p\tilde{S}\big[\lambda(1-z)\big] + (1-p)w\tilde{U}\big[\lambda(1-z)\big]$$

as follows. Define the 'descendants' of a customer to be the new arrivals during his service time. Then the number of descendants of a 'new' customer has generating function $p\tilde{S}\big[\lambda(1-z)\big]+(1-p)w\tilde{U}\big[\lambda(1-z)\big]$. The extra factor w indicates the fact that, with probability $1-p$, the 'new' customer also generates one 'old' customer, which is he himself, being fed back. In the same way one can easily handle the case of M different customer types. Formula (2.5) would keep the same structure, the $(k+1)$th element of G in the righthand side becoming (with an obvious extension of notation): $p_k\tilde{S}_k(\lambda(1-z)) \ + \ (1-p_k)w_{k+1}\tilde{U}_k(\lambda(1-z))$, $k = 1, \ldots, M-2$.
We now solve (2.5) by iteration. For $|z| \leq 1, |w| \leq 1$, define

$$\begin{aligned} &m_N^{(0)}(z,w) := z, \quad m_O^{(0)}(z,w) := w, \\ &m_N^{(1)}(z,w) := p\tilde{S}\big[\lambda(1-z)\big] + (1-p)w\tilde{U}\big[\lambda(1-z)\big], \\ &m_O^{(1)}(z,w) := p_1\tilde{S}_1\big[\lambda(1-z)\big] + (1-p_1)w\tilde{U}_1\big[\lambda(1-z)\big] \ . \end{aligned} \tag{2.6}$$

For $n \geq 1$,

$$\begin{aligned} &m_N^{(n)}(z,w) = p\tilde{S}\Big[\lambda\big(1-m_N^{(n-1)}(z,w)\big)\Big] \\ + \quad &(1-p)m_O^{(n-1)}(z,w)\tilde{U}\Big[\lambda\big(1-m_N^{(n-1)}(z,w)\big)\Big], \\ &m_O^{(n)}(z,w) = p_1\tilde{S}_1\Big[\lambda\big(1-m_N^{(n-1)}(z,w)\big)\Big] \\ + \quad &(1-p_1)m_O^{(n-1)}(z,w)\tilde{U}_1\Big[\lambda\big(1-m_N^{(n-1)}(z,w)\big)\Big] \ . \end{aligned} \tag{2.7}$$

Note that $m_N^{(n)}(z,w)$ (respectively, $m_O^{(n)}(z,w)$) may be interpreted as the nth generation descendants of one new (respectively, old) customer. Formula (2.5) can be rewritten as

$$G(z,w) = \tilde{V}(\lambda(1-z))G\left(m_N^{(1)}(z,w), m_O^{(1)}(z,w)\right). \tag{2.8}$$

Iterating this formula k times results in

$$G(z,w) = \prod_{n=0}^{k} \tilde{V}\left[\lambda\left(1 - m_N^{(n)}(z,w)\right)\right] \cdot G\left(m_N^{(k+1)}(z,w), m_O^{(k+1)}(z,w)\right) .$$

It can be shown that, if $\rho < 1$, $\lim_{n\to\infty} m_N^{(n)}(z,w) = 1$ and $\lim_{n\to\infty} m_O^{(n)}(z,w) = 1$; convergence is geometrically fast, so that $\prod_{n=0}^{\infty} \tilde{V}\left[\lambda\left(1 - m_N^{(n)}(z,w)\right)\right]$ is a convergent infinite product. Hence, for $\rho < 1$, as $G(1,1) = 1$,

$$G(z,w) = \prod_{n=0}^{\infty} \tilde{V}\left[\lambda\left(1 - m_N^{(n)}(z,w)\right)\right], \quad |z| \le 1, \ |w| \le 1. \tag{2.9}$$

Moments

¿From (2.2), using the Poisson arrival property, we readily obtain

$$\begin{aligned} E[N] &= \lambda\{b_N E[N] + b_O E[O] + v\}, \\ E[O] &= (1-p)E[N] + (1-p_1)E[O]. \end{aligned} \tag{2.10}$$

Thus,

$$E[O] = \frac{1-p}{p_1} E[N], \tag{2.11}$$

and hence, using (1.1),

$$E[N] = \rho E[N] + \lambda v. \tag{2.12}$$

This implies

$$E[N] = \frac{\lambda v}{1-\rho}. \tag{2.13}$$

Clearly, result (2.12) that leads to (2.13) can be directly obtained by using balance arguments.

Finally, the mean *cycle time*, i.e., the time between two consecutive vacation beginnings, is given by

$$E[C] = E[N]/\lambda = \frac{v}{1-\rho}. \tag{2.14}$$

3 The steady-state joint queue length distribution

$G(z,w)$ is the PGF of the steady-state joint queue length distribution at the *beginning* of a session (end of a vacation). Let $G^{(e)}(z,w)$ denote the corresponding PGF at the *end* of a session (beginning of a vacation). Also, let $H_N(z,w)$ and $H_O(z,w)$ be the joint queue length PGF at the *beginning* of a service attempt of a new (respectively, old) customer. Similarly, let $H_N^{(e)}(z,w)$ and $H_O^{(e)}(z,w)$ be the corresponding PGF's right after the *end* of a service. Considering one cycle (service plus vacation), the PGF of the state of the system at an arbitrary moment, $F^*(z,w)$, is given by

$$F^*(z,w) = \frac{1}{E[C]}\Big[b_N E[N] F^*\Big(z,w\Big|\begin{matrix}\text{a new customer}\\ \text{is served}\end{matrix}\Big) \tag{3.1}$$
$$+\, b_O E[O] F^*\Big(z,w\Big|\begin{matrix}\text{an old customer}\\ \text{is served}\end{matrix}\Big) + vF^*\Big(z,w\Big|\begin{matrix}\text{the server is}\\ \text{on vacation}\end{matrix}\Big)\Big]\ .$$

Now,

$$F^*\Big(z,w\Big|\begin{matrix}\text{the server is}\\ \text{on vacation}\end{matrix}\Big) = G^{(e)}(z,w)\cdot\frac{1-\tilde{V}\big[\lambda(1-z)\big]}{v\big(\lambda(1-z)\big)}, \tag{3.2}$$

where $\frac{1-\tilde{V}(s)}{vs}$ is the LST of the past part, V_P, of V;

$$F^*(z,w \mid \text{a new customer is served}) = H_N(z,w)\cdot\frac{1-\tilde{B}_N\big[\lambda(1-z)\big]}{b_N\big(\lambda(1-z)\big)}, \tag{3.3}$$

$$F^*(z,w \mid \text{an old customer is served}) = H_O(z,w)\cdot\frac{1-\tilde{B}_O\big[\lambda(1-z)\big]}{b_O\big(\lambda(1-z)\big)}. \tag{3.4}$$

Substituting (3.2), (3.3) and (3.4) in (3.1), and then using

$$G(z,w) = G^{(e)}(z,w)\tilde{V}\big[\lambda(1-z)\big], \tag{3.5}$$

we obtain

$$\begin{aligned} F^*(z,w) &= \frac{1}{E[C]\lambda(1-z)}[[E[N]\Big(1-\tilde{B}_N\big[\lambda(1-z)\big]\Big)H_N(z,w) \\ &+ E[O]\Big(1-\tilde{B}_O\big[\lambda(1-z)\big]\Big)H_O(z,w)] \\ &\quad +\Big(1-\tilde{V}\big[\lambda(1-z)\big]\Big)G(z,w)/\tilde{V}\big[\lambda(1-z)\big]]. \end{aligned} \tag{3.6}$$

The steady-state joint queue length distribution of old and new customers depends on the order in which customers are served during a session. We now assume that during each session of the server, new customers have priority over old ones, that is, they are served first. Thus, if $G(z,w)$ is the PGF of the number of customers at the beginning of a session, let $G_N^{(e)}(z,w)$ be the PGF of the number of customers at the moment where the server ends serving *new* customers. We now use a nice idea that seems to be due to Eisenberg [5] in a polling context, see also Altman and Yechiali [1], Borst and Boxma [2], and Sidi et al. [6]. It is the observation that the union of all session beginnings and all service endings of new customers coincides with the union of all service beginnings of new customers and endings of whole service periods of new customers. In terms of generating functions:

$$G(z,w) + E[N]H_N^{(e)}(z,w) = G_N^{(e)}(z,w) + E[N]H_N(z,w). \tag{3.7}$$

Similarly, with $G_O(z,w)$ denoting the PGF of the number of customers at the moment where the server starts serving *old* customers:

$$G_O(z,w) + E[O]H_O^{(e)}(z,w) = G^{(e)}(z,w) + E[O]H_O(z,w). \tag{3.8}$$

Above we might as well have written $G_N^{(e)}(z,w)$ instead of $G_O(z,w)$; note that during a session the period in which the server serves new (or old) customers may have length zero, but this does not cause complications. Now

$$G_O(z,w) = G_N^{(e)}(z,w) = E\big[(m_N^{(1)}(z,w))^N w^O\big] = G(m_N^{(1)}(z,w),w). \tag{3.9}$$

Using (3.5), (3.9) and the obvious relations

$$H_N^{(e)}(z,w) = H_N(z,w)m_N^{(1)}(z,w)/z, \tag{3.10}$$

$$H_O^{(e)}(z,w) = H_O(z,w)m_O^{(1)}(z,w)/w, \tag{3.11}$$

equations (3.7) and (3.8) are written as

$$E[N]H_N(z,w)[1 - m_N^{(1)}(z,w)/z] = G(z,w) - G(m_N^{(1)}(z,w),w), \tag{3.12}$$

$$E[O]H_O(z,w)[1-m_O^{(1)}(z,w)/w] = G(m_N^{(1)}(z,w),w) - G(z,w)/\tilde{V}(\lambda(1-z)). \tag{3.13}$$

Substituting $E[N]H_N(z,w)$ from (3.12) and $E[O]H_O(z,w)$ from (3.13) in equation (3.6) gives the result we were looking for: the PGF $F^*(z,w)$ of the joint steady-state queue length distribution, expressed in terms of the known $G(\cdot,\cdot)$.

Briefly consider the case where old customers have priority over new ones. Let $G_O^{(e)}(z,w) = G_N(z,w)$ be the PGF of the joint queue length distribution at the moment where the server ends serving old customers and starts serving new customers. Then (3.7) and (3.8) are modified into

$$\begin{aligned} G(z,w) + E[O]H_O^{(e)}(z,w) &= G_O^{(e)}(z,w) + E[O]H_O(z,w), && (3.14) \\ G_N(z,w) + E[N]H_N^{(e)}(z,w) &= G^{(e)}(z,w) + E[N]H_N(z,w). && (3.15) \end{aligned}$$

Instead of (3.9) we have

$$G_O^{(e)}(z,w) = G_N(z,w) = E\big[z^N (m_O^{(1)}(z,w))^O\big] = G(z, m_O^{(1)}(z,w)). \tag{3.16}$$

Similar manipulations as above and finally substitution in (3.6) gives $F^*(\cdot,\cdot)$ in terms of $G(\cdot,\cdot)$.

4 Waiting Times

The waiting time distribution obviously depends on the order of service. We assume that the order of service follows the "longest present first" rule. That is, old customers are served before new customers and within each group the order is FCFS. We shall express the LST of the waiting time into the just obtained function $G(\cdot,\cdot)$.
For the calculation of the LST of the waiting time, W, of an arbitrary customer, we consider here the case where $p_1 = 1$. That is, a customer is either successful on his first service attempt or, if fed back after the unsuccessful service, his second service attempt is *always* successful (having duration S_1). More involved cases can in principle be handled in a similar manner, but the analysis may become quite complicated.
Let W_S (W_U) denote the waiting time of a customer who is successful (unsuccessful) on his initial service attempt. Then,

$$P(W \leq t) = pP(W_S \leq t) + (1-p)P(W_U \leq t)\ . \tag{4.1}$$

Calculation of the distribution of W_S
Consider a tagged successful customer K. We can write $W_S = C_R + D_1 + D_2$. Here C_R denotes the residual cycle time, measured from the moment of the arrival of K. D_1 is the sum of the service times of those $F(N)$ customers, out of N present at the *start* of the cycle, who were fed back during the cycle (the old customers were successful and left the system). Each of those customers now receives a successful service S_1. D_2 is the sum of service times of *new* arrivals during the *past* part, C_P, of the cycle (before the arrival of K). Each of those new customers requires a service duration B_N. Thus,

$$\begin{aligned} E\left[e^{-\omega D_2} \mid C_P = t\right] &= E\Big[E\left[e^{-\omega D_2} \mid A(t)\right]\Big| C_P = t\Big] \\ &= E\left[\left(\tilde{B}_N(\omega)\right)^{A(t)} \mid C_P = t\right] = e^{-\lambda\left(1-\tilde{B}_N(\omega)\right)t}, \end{aligned} \tag{4.2}$$

where $\tilde{B}_N(\omega) = p\tilde{S}(\omega) + (1-p)\tilde{U}(\omega)$, and $A(t)$ is the number of Poisson arrivals during a time interval of length t.
Also,

$$E\left[e^{-\omega D_1} | F(N)\right] = \left[\tilde{S}_1(\omega)\right]^{F(N)}. \tag{4.3}$$

Combining the above, we have

$$\begin{aligned} E[e^{-\omega W_S}] = \int_{t_P=0}^{\infty}\int_{t_R=0}^{\infty} dP\{C_P \leq t_P, C_R \leq t_R\} e^{-\omega t_R} e^{-\lambda\left(1-\tilde{B}_N(\omega)\right)t_P}. \\ \cdot \sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\sum_{\ell=0}^{k} P(N=k, O=j, F(N)=k-\ell \mid C_P = t_P, C_R = t_R)\tilde{S}_1(\omega)^{k-\ell}\ . \end{aligned} \tag{4.4}$$

It follows from renewal theory (cf. Formula (I.6.23) of [4]) that one can write:

$$\int_{t_P=0}^{\infty}\int_{t_R=0}^{\infty} g(t_P,t_R)dP\{C_P\le t_P, C_R\le t_R\}$$
$$= \int_{x=0}^{\infty}\int_{t_R=0}^{x} g(x-t_R,t_R)\frac{dt_R}{x}\frac{xdP(C\le x)}{E[C]}\ .$$

Hence,

$$E[e^{-\omega W_S}] = \tfrac{1}{E[C]}\int_{x=0}^{\infty}\int_{t_R=0}^{x} dP(C\le x)e^{-\omega t_R}e^{-\lambda(1-\tilde{B}_N(\omega))(x-t_R)}dt_R\ \cdot$$
$$\cdot \sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\sum_{\ell=0}^{k} P(N=k,O=j,F(N)=k-\ell\mid C=x)\tilde{S}_1(\omega)^{k-\ell}\quad(4.5)$$

Using $P\big(F(N)=k-\ell\mid N=k,O=j\big)=\binom{k}{\ell}p^\ell(1-p)^{k-\ell}$ and performing the integration in (4.5) with respect to t_R we have, letting $\xi(\omega)\equiv\lambda\big(1-\tilde{B}_N(\omega)\big)$,

$$E[e^{-\omega W_S}] = \frac{1}{E[C]}\frac{1}{\omega-\xi(\omega)}\int_{x=0}^{\infty}(e^{-\xi(\omega)x}-e^{-\omega x})\ \cdot$$
$$\cdot\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\sum_{\ell=0}^{k} P(N=k,O=j)\binom{k}{\ell}p^\ell(1-p)^{k-\ell}\tilde{S}_1(\omega)^{k-\ell}\ \cdot$$
$$\cdot dP(C\le x\mid N=k,O=j,F(N)=k-\ell)$$
$$= \frac{1}{E[C]}\frac{1}{\omega-\xi(\omega)}\int_{x=0}^{\infty}(e^{-\xi(\omega)x}-e^{-\omega x})\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\sum_{\ell=0}^{k} P(N=k,O=j)\ \cdot$$
$$\cdot\tilde{S}_1(\omega)^{k-\ell}\binom{k}{\ell}p^\ell(1-p)^{k-\ell}d\big[S^{\ell *}*U^{(k-\ell)*}*S_1^{j*}*V\big](x)$$
$$= \frac{1}{E[C]}\frac{1}{\omega-\xi(\omega)}\sum_{k=0}^{\infty}\sum_{j=0}^{\infty} P(N=k,O=j)\ \cdot$$
$$\cdot\sum_{\ell=0}^{k}\Bigg\{\Big[\tilde{S}\big(\xi(\omega)\big)\Big]^\ell\Big[\tilde{U}\big(\xi(\omega)\big)\Big]^{k-\ell}\Big[\tilde{S}_1\big(\xi(\omega)\big)\Big]^j\tilde{V}\big(\xi(\omega)\big)$$
$$-\big[\tilde{S}(\omega)\big]^\ell\big[\tilde{U}(\omega)\big]^{k-\ell}\big[\tilde{S}_1(\omega)\big]^j\tilde{V}(\omega)\Bigg\}\binom{k}{\ell}p^\ell(1-p)^{k-\ell}\tilde{S}_1(\omega)^{k-\ell}\ .$$

Thus,

$$E\big[e^{-\omega W_S}\big] = \frac{1}{E[C]}\frac{1}{\omega-\xi(\omega)}\ \cdot \qquad (4.6)$$
$$\cdot\Bigg[\tilde{V}\big(\xi(\omega)\big)G\Big(p\tilde{S}\big(\xi(\omega)\big)+(1-p)\tilde{U}\big(\xi(\omega)\big)\tilde{S}_1(\omega),\tilde{S}_1\big(\xi(\omega)\big)\Big)$$
$$-\tilde{V}(\omega)G\big(p\tilde{S}(\omega)+(1-p)\tilde{U}(\omega)\tilde{S}_1(\omega),\tilde{S}_1(\omega)\big)\Bigg]\ .$$

Remark. For $p = 1$, i.e., the case without feedback, the expression in (4.6) reduces to the LST of the waiting distribution in the M/G/1 queue with multiple gated vacations, cf. [8], Formula (5.21b) on p. 208. The latter formula can be decomposed into the waiting time LST of the ordinary M/G/1 queue and an additional term that is due to the occurrence of vacations. In (4.6) one also sees a hint of a decomposition; observe that the waiting time LST in the M/G/1 queue with service time LST $\tilde{B}_N(\omega)$ is given by $[1 - \lambda(ps + (1-p)u)]\omega/[\omega - \xi(\omega)]$.

Calculation of the distribution of W_U
Consider the waiting time of an unsuccessful tagged customer K, arriving during a cycle C. Observe that, because old customers are served before new ones in a session, the first service time of K does not influence his waiting time between both services; hence we can obtain his sojourn time distribution by convoluting his waiting time distribution with his two service time distributions.
We have

$$\begin{aligned} W_U = C_R \quad &+ \quad \text{service of } F(N) \text{ unlucky customers who were} \\ & \quad \text{present at the beginning of the cycle } C \\ &+ \quad \text{service of all arrivals in } C_P \text{ (possibly two services)} \\ &+ \quad \text{vacation at end of 2nd cycle} \\ &+ \quad \text{first service of all (new) arrivals in } C_R \ . \end{aligned} \tag{4.7}$$

Thus,

$$\begin{aligned} E\big[e^{-\omega W_U}\big] &= \tilde{V}(\omega)\frac{1}{E[C]}\int_{x=0}^{\infty}\int_{t_R=0}^{x} e^{-\omega t_R} dP(C \le x) dt_R \cdot \\ &\cdot \sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\sum_{\ell=0}^{k} P(N = k, O = j, F(N) = k - \ell \mid C = x)\tilde{S}_1(\omega)^{k-\ell} \cdot \\ &\cdot \sum_{m=0}^{\infty} e^{-\lambda(x-t_R)}\frac{[\lambda(x-t_R)]^m}{m!}\sum_{n=0}^{\infty} e^{-\lambda t_R}\frac{(\lambda t_R)^n}{n!} \cdot \\ &\cdot \sum_{i=0}^{m}\binom{m}{i} p^i(1-p)^{m-i}\tilde{S}(\omega)^i\big(\tilde{U}(\omega)\tilde{S}_1(\omega)\big)^{m-i} \cdot \\ &\cdot \sum_{r=0}^{n}\binom{n}{r} p^r(1-p)^{n-r}\tilde{S}(\omega)^r\tilde{U}(\omega)^{n-r} \ . \end{aligned} \tag{4.8}$$

Setting $\tilde{B}_2(\omega) = p\tilde{S}(\omega) + (1-p)\tilde{U}(\omega)\tilde{S}_1(\omega)$ we have

$$E\big[e^{-\omega W_U}\big] = \tilde{V}(\omega)\frac{1}{E[C]}\int_{x=0}^{\infty}\int_{t_R=0}^{x} e^{-\omega t_R} dt_R \cdot \tag{4.9}$$

$$\cdot \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \sum_{\ell=0}^{\infty} dP(C \leq x \mid N=k, O=j, F(N)=k-\ell) P(N=k, O=j) \cdot$$

$$\cdot P\big(F(N)=k-\ell \mid N=k, O=j\big) \tilde{S}_1(\omega)^{k-\ell} e^{-\lambda t_R[1-\tilde{B}_N(\omega)]} e^{-\lambda(x-t_R)[1-\tilde{B}_2(\omega)]}$$

$$= \tilde{V}(\omega) \frac{1}{E[C]} \int_{x=0}^{\infty} \int_{t_R=0}^{x} e^{-\omega t_R} dt_R \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \sum_{\ell=0}^{k} d\big[S^{\ell *} * U^{(k-\ell)*} * S_1^{j*} * V\big](x) \cdot$$

$$\cdot P(N=k, O=j) \binom{k}{\ell} p^{\ell}(1-p)^{k-\ell} \tilde{S}_1(\omega)^{k-\ell} e^{-\lambda t_R[1-\tilde{B}_N(\omega)]} e^{-\lambda(x-t_R)[1-\tilde{B}_2(\omega)]}.$$

Hence, integrating with respect to t_R, we get

$$E\big[e^{-\omega W_U}\big] = \frac{\tilde{V}(\omega)}{\omega + \lambda(\tilde{B}_2(\omega) - \tilde{B}_N(\omega))} \frac{1}{E[C]} \int_{x=0}^{\infty} \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} P(N=k, O=j) \cdot$$

$$\cdot \sum_{\ell=0}^{k} \binom{k}{\ell} p^{\ell}(1-p)^{k-\ell} \tilde{S}_1(\omega)^{k-\ell} e^{-\lambda x[1-\tilde{B}_2(\omega)]} \Big[1 - e^{-x\left(\omega + \lambda(\tilde{B}_2(\omega) - \tilde{B}_N(\omega))\right)}\Big] \cdot$$

$$\cdot d\big[S^{\ell *} * U^{(k-\ell)*} * S_1^{j*} * V\big](x)$$

$$= \frac{\tilde{V}(\omega)}{\omega + \lambda\big(\tilde{B}_2(\omega) - \tilde{B}_N(\omega)\big)} \frac{1}{E[C]} \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} P(N=k, O=j) \cdot$$

$$\cdot \sum_{\ell=0}^{k} \binom{k}{\ell} p^{\ell}(1-p)^{k-\ell} \tilde{S}_1(\omega)^{k-\ell} \Big\{ \tilde{S}\big(\xi_2(\omega)\big)^{\ell} \tilde{U}\big(\xi_2(\omega)\big)^{k-\ell} \tilde{S}_1\big(\xi_2(\omega)\big)^{j} \tilde{V}\big(\xi_2(\omega)\big)$$

$$- \tilde{S}\big(\omega + \xi(\omega)\big)^{\ell} \tilde{U}\big(\omega + \xi(\omega)\big)^{k-\ell} \tilde{S}_1\big(\omega + \xi(\omega)\big)^{j} \tilde{V}\big(\omega + \xi(\omega)\big) \Big\},$$

where $\xi_2(\omega) \equiv \lambda\big(1 - \tilde{B}_2(\omega)\big)$.
Finally,

$$E\big[e^{-\omega W_U}\big] = \frac{\tilde{V}(\omega)}{\omega + \lambda\big(\tilde{B}_2(\omega) - \tilde{B}_N(\omega)\big)} \frac{1}{E[C]} \cdot \tag{4.10}$$

$$\cdot \Bigg[\tilde{V}\big(\xi_2(\omega)\big) G\Big(p\tilde{S}\big(\xi_2(\omega)\big) + (1-p)\tilde{U}\big(\xi_2(\omega)\big)\tilde{S}_1(\omega),\ \tilde{S}_1\big(\xi_2(\omega)\big)\Big)$$

$$- \tilde{V}\big(\omega + \xi(\omega)\big) G\Big(p\tilde{S}\big(\omega + \xi(\omega)\big) + (1-p)\tilde{U}\big(\omega + \xi(\omega)\big)\tilde{S}_1(\omega),\ \tilde{S}_1\big(\omega + \xi(\omega)\big)\Big) \Bigg].$$

The waiting time LST of an arbitrary customer follows from (4.1), (4.6) and (4.10). In particular, moments can be obtained straightforwardly.

Acknowledgment. The first author's research has been supported by the European Grant BRA-QMIPS of CEC DG XIII.

5 References

[1] E. Altman, U. Yechiali (1994). Polling in a closed network. *Probability in the Engineering and Informational Sciences* 8, 327-343.

[2] S.C. Borst, O.J. Boxma (1994). Polling models with and without switchover times. CWI Report BS-R9421; to appear in *Operations Research*.

[3] O.J. Boxma (1994). Polling systems. In: K.R. Apt, A. Schrijver and N.M. Temme (eds.), *From Universal Morphisms to Megabytes: A Baayen Space Odyssey* (Liber Amicorum for P.C. Baayen), pp. 215-230. CWI, Amsterdam, December 1994.

[4] J.W. Cohen (1982). *The Single Server Queue.* North-Holland Publ. Cy., Amsterdam (2nd ed.).

[5] M. Eisenberg (1972). Queues with periodic service and changeover time. *Operations Research* 20, 440-451.

[6] M. Sidi, H. Levy and S.W. Fuhrmann (1992). A queueing network with a single cyclically roving server. *Queueing Systems* 11, 121-144.

[7] L. Takács (1963). A queueing model with feedback. *R.A.I.R.O. Recherche Opérationelle/Operations Research* 11, 345-354.

[8] H. Takagi (1991). *Queueing Analysis, Vol. 1: Vacation and Priority Systems.* North-Holland Publ. Cy., Amsterdam.

[9] T. Takine, H. Takagi and T. Hasegawa (1991). Sojourn times in vacation and polling systems with Bernoulli feedback. *J. Appl. Probab.* 28, 422-432.