MPE — Volume 2, pp. 95-106
Reprints available directly from the publisher
Photocopying permitted by license only

BULK INPUT QUEUES WITH QUORUM AND MULTIPLE VACATIONS

JEWGENI H. DSHALALOW and JAY YELLEN

Applied Mathematics Program, Florida Institute of Technology, Melbourne, FL 32901, USA

(Received April 17, 1995)

The authors study a single-server queueing system with bulk arrivals and batch service in accordance to the general quorum discipline: a batch taken for service is not less than r and not greater than R ($\geq r$). The server takes vacations each time the queue level falls below r (≥ 1) in accordance with the multiple vacation discipline. The input to the system is assumed to be a compound Poisson process. The analysis of the system is based on the theory of first excess processes developed by the first author. A preliminary analysis of such processes enabled the authors to obtain all major characteristics for the queueing process in an analytically tractable form. Some examples and applications are given.

KEYWORDS: *Queueing process*; *first excess level process*; *first passage time*; *termination index*; *quorum*; *vacations*; *multiple vacations*; *bulk input*; *batch service*; *fluctuation theory*; *embedded Markov chain*; *equilibrium*
AMS Nos. 60K10, 60K15, 60K25.

1. INTRODUCTION

In this paper we study a general M/G/1-type queue with multiple vacations and nonexhaustive service. A basic M/G/1-vacation queue refers to a queueing system with a Poisson input stream of customers, generally distributed service times, and a single server processing customers one at a time. When the inventory of available customers in the system is exhausted, the server, instead of waiting for new customers, leaves the system for vacation. Systems with vacationing servers were first introduced in the sixties. However, they were not regarded practical enough and were largely ignored. In the seventies and eighties they were referred as *T-policy queues* [1] or queues with *continuously operating server* [2]. The latter name was an apt description. Indeed, the server pretended to process a phantom customer, and did not interrupt its service even when new customers arrived and waited. Only in the seventies, did systems with "nonwaiting" servers become more accepted and gained in popularity.

One came to realize that there were practical situations in which the server does not wait. He may arrange some preventive maintenance or even physically leave the system to perform other jobs of second priority. It was logical to assume that they should not be interrupted if during such work new customers arrive. The maintenance jobs could be adjusted in such a way that their durations are distributed differently from the usual service times. There are also a large variety of queues with priority customers and server

breakdowns which, in recent surveys on vacation models (*cf.* Doshi [3]), may be interpreted as systems with vacations. However, these are preemptive and probably should not be included in an already broad class of "true" vacation systems with nonpreemptive disciplines.

During the eighties, other variants of vacation policies were introduced and explored including those with single or multiple vacations. In the first case, the server leaves the system when the queue becomes empty for a single vacation segment. Thus, if during his vacation no customer arrives at the system, he assumes the policy of a waiting server. In the second case, the server starts a vacation consisting of multiple segments. At the end of the nth segment, the server returns to the system to check its status, and, if no customer is present, he begins his $(n + 1)$st vacation segment, and so on. In both cases, the policy of starting a vacation is referred to as *exhaustive*.

A more general scenario is one in which the server leaves on vacation when the queue drops below some fixed level, say, $r \geq 1$. It would also make sense if the server takes more than one customer for service at a time. In addition, it would make sense to interrupt the server's sequence of vacation segments when the queue returns to its "critical" length of at least r customers. This policy is referred to as *nonexhaustive*, and more specifically, as *quorum*. (Note that systems on quorum do not necessarily assume vacationing servers. The same policy of exiting and entering busy periods can be applied to a "dormant" server, see for example [4] and [5].) The analysis becomes more difficult if one allows the input to be bulk, since in this case it is not easy to determine when (*first passage time*) the queue hits (i.e., reaches or exceeds) that critical level r, and what its size (*excess level*) would then be. Although a few papers [6–8] did consider this general scenario, their works did not analyze first passage times and first excess levels in its general form. The first result on first excess level processes appeared in 1992 [9] which made it possible to analyze a broad class of bulk input quorum systems with or without vacations [4, 10, 11, 12]. Another paper with vacationing server appeared soon thereafter in Dshalalow [13], where the first excess level analysis was applied to an r-quorum system.

In the present paper, we further generalize such a system by allowing a more flexible service discipline. In this system, the server starts a multiple vacation (i.e., a sequence of vacation segments) whenever the queue drops below level r, and it resumes service at the end of a vacation segment when the queue accumulates to at least r customers. Here we assume that the server capacity is $R \geq r$, and that the server takes any number of customers available between r and R. We also incorporate the latest results on first excess level theory and present some important special cases and applications.

The paper is organized as follows. In Section 2 we formally define the system's arrival, service and vacation disciplines. The system under study is one with bulk Poisson arrival flow, generally distributed service and vacation times, (r,R)-quorum service discipline, and multiple nonexhaustive vacations. It is also assumed that distributions of service and vacation segments are different. In Section 3 we consider a process embedded in the queueing process over the times of service completions. It is closely related to another process embedded in the queueing process over the first passage times. An explicit relationship and a necessary and sufficient condition for the ergodicity of the embedded process are established. Under this condition, the probability generating function of the limiting distribution of the process, obtained in an analytically tractable form, appears in Section 6. In Section 4 we give all pertinent results on the first excess level theory applied

to our system, and in Section 5, all related preliminaries. Section 7 concludes with some applications (conservation law of the system).

2. NOTATION AND PRELIMINARIES

The system involves three different point processes on the time axis, representing times of arrivals (of batches), times of service completions, and ending times of vacation segments. We will refer to these three processes by using τ^a, τ^s, and τ^v, respectively.

Arrivals Discipline

τ_k^a will denote the arrival time of the kth batch of customers and X_k its size. Let $A_n = \sum_{k=1}^{n} X_k$. Therefore, A_n is the total number of customers arrived at the system in the time interval $[0, \tau_n^a]$. We assume that $(\tau^a, A) = \sum_{n=1}^{\infty} X_n \epsilon_{\tau_n^a}$ is a compound Poisson process with intensity λ obtained from position independent marking, where ϵ_a is the point mass concentrated at a. We also assume that the set of marks $A = \{A_1, A_2, \ldots\}$ forms a discrete-valued renewal process. For all X_k we denote by $a(u)$ their common probability generating function (PGF) $\mathbf{E}[u^{X_k}]$, $k = 1,2,\ldots$, with mean α.

Service Discipline

We assume that the server capacity is R and the server takes any group of customers for service whose size is between r and R inclusive, $r \leq R$. If the queue is less than r, then the server leaves the system ("goes on vacation") until the waiting room has at least r customers. This service discipline, referred to as "*(r,R)-quorum*," was introduced by M. Neuts [5] in 1967, and later surveyed by Chaudhry and Templeton in their monograph [2] in a number of special cases. The practicality of this discipline attracted many authors since, such as [4, 14, 6–8, 11, 12]. τ_n^s will denote the time of the nth service completion, $n = 0,1,\ldots$, and δ_n^s its duration for $n \geq 1$. τ_0^s will be taken to be zero. Assume that δ_n^s's are independent, identically distributed random variables that are independent of (τ^a, A) and let $F^s(t) = \mathbf{P}\{\delta_1^s \leq t\}$ be their common distribution function (DF) with finite mean μ^s.

Vacation Discipline

If, upon service completion, the queue size is less than r, the server leaves on vacation. Each vacation consists of a sequence of segments whose durations are assumed to be independent and identically distributed random variables. We also assume that the distributions associated with any two vacations are identical. Our analysis will not require us to distinguish between two different vacations, and to simplify our notation, we will suppress the index referencing a particular vacation. Thus, τ_k^v will denote the time that the kth segment of a particular vacation ends and δ_k^v will be its duration with $F^v(t) = \mathbf{P}\{\delta_1^v \leq t\}$ and finite mean μ^v. Let Δ be the random variable representing the number of segments of that vacation. We assume that the server will be required to take Δ vacation segments if at the end of the Δth segment the queue length hits (i.e., reaches or exceeds)

level r for the first time. Observe that no particular vacation segment is interrupted should the queue hit r during that segment. The described vacation discipline for $r = 1$ is referred to as a *non-preemptive multiple vacation discipline*. To distinguish this standard notion from ours we add "*r-quorum*" to this name. Although this addition appears to resemble a well-known "N-policy" discipline (combined with vacations), the two systems are different. A basic N-policy discipline is one where the server takes exactly one customer at a time for service, and leaves on vacation (or, alternatively, idles) whenever the queue is empty. The sequence of vacation segments is terminated when the queue hits level N. Such a system was considered by [7, 8, 11, 12]. A generalized discipline with r-quorum and N-policy would thus be one where r and N are generally different ($r \leq N$). (See Muh [11, 12].) In our system, $r = N$, but it includes vacation segments.

3. EMBEDDED QUEUEING PROCESS

Let $Q(t)$ be the number of customers in the system at time t including those in service, and assume that $Q(t)$ is a process with *a.s.* right-continuous paths. Let Q_n^s and Q_k^v denote $Q(\tau_n^s)$ and $Q(\tau_k^v)$, respectively. By definition of Δ, we have $Q_k^v < r$ for $1 \leq k < \Delta$, and $Q_\Delta^v \geq r$.

Let X_n^s denote the number of arrivals during the nth service and let X_k^v be the number of arrivals during a particular vacation's kth segment. Then,

$$\mathbf{E}[u^{X_n^s}] = \int_0^\infty e^{\lambda t(a(u)-1)} dF^s(t) \text{ and } \mathbf{E}[u^{X_k^v}] = \int_0^\infty e^{\lambda t(a(u)-1)} dF^v(t).$$

We denote these two transforms by $\varphi^s(u)$ and $\varphi^v(u)$, respectively.

We will call the time interval between two successive service completions a *service cycle*. The first service cycle starts at time $t = 0$. If the queue length at zero, Q_0^s, is at least r, then the server takes a group of customers for service not exceeding its capacity R and the first service ends at time τ_1^s. In this case, the length of the first service cycle coincides with the service duration δ_1^s (i.e., $\tau_1^s = \delta_1^s$). If Q_0^s is less than r, the server goes on vacation, which consists of Δ segments, at which time the queue accumulates to Q_Δ^v customers. In this case, the service cycle consists of the vacation interval $[0, \tau_\Delta^v]$, where τ_Δ^v is the so-called *first passage time*, and the interval of service $(\tau_\Delta^v, \tau_\Delta^v + \delta_1^s]$. Thus, if we denote the queue length at the beginning of the first service by Q_1^{bs}, then

$$Q_1^{bs} = \begin{cases} Q_\Delta^v, & Q_0^s < r \\ Q_0^s, & Q_0^s \geq r. \end{cases} \tag{3.1}$$

In light of the above notation, our service and vacation disciplines imply that

$$Q_1^s = Q_1^{res} + X_1^v, \tag{3.2}$$

where

$$Q_1^{res} = (Q_1^{bs} - R)^+ \tag{3.3}$$

($y^+ \triangleq \max\{y,0\}$) represents the rest of the customers waiting for next service at the beginning of the first service period. Relation (3.2) defines recursively the sequence $Q^s = \{Q_n^s; n = 0,1,\ldots\}$ of random variables which under current assumptions imposed on the input stream forms a homogeneous Markov chain (MC). Let P be its transition probability matrix with $P_i(u)$ being the PGF of the ith row. Then, $P_i(u) = \mathbf{E}^i[u^{Q_1^s}]$, where $\mathbf{E}^i$ is the expectation with respect to the probability measure $\mathbf{P}^i = \mathbf{P}(\cdot|Q_0^s = i)$.

Using a sequence of standard arguments one can confirm that P is a $\Delta_{R,R}$-type stochastic matrix which satisfies necessary and sufficient conditions established by Abolnikov and Dukhovny to insure that the process Q^s is ergodic. The reader is referred to articles [4, 13, 10] where such analysis is done in great detail. According to Abolnikov and Dukhovny criterion, the MC Q^s is ergodic if and only if

$$\rho^s \triangleq \lambda\alpha\mu^s < R, \tag{3.4}$$

where λ is the mean number of arriving batches per unit of time, α is the average size of an arriving batch, and μ^s is the expected duration of a service period. The value ρ^s represents the average number of arrivals per service time. The ergodicity condition states that this number should necessarily be strictly less than server capacity. Under this condition, the stationary distribution of Q^s exists and coincides with the unique normalized fixed point, p, of the operator P (known as the "invariant probability measure"):

$$\boldsymbol{p} = \boldsymbol{p}^P, \boldsymbol{p} > 0 \text{ (vector), and } 1 = \boldsymbol{p}\mathbf{1} \text{ (scalar product of } \boldsymbol{p} = (p_i;\, i \geq 0)^{\mathrm{T}} \text{ and vector } \mathbf{1}). \tag{3.5}$$

The latter can routinely be shown to yield the equivalent functional equation

$$P(u) = \sum_{i=0}^{\infty} P_i(u)p_i,\ P(1) = 1, \tag{3.6}$$

where $P(u)$ is the z-transform of vector $\boldsymbol{p}$ and p_i is its ith element. The evaluation of $P_i(u)$ is crucial to obtain $P(u)$ in an analytically tractable form. The next section is devoted to a special case of the first excess level theory initiated in Abolnikov and Dshalalow [9] and further developed in a few subsequent articles [10, 15–17]. This treatment involves $P_i(u)$ after which we return to the probability generating function $P(u)$ of the stationary distribution $\boldsymbol{p}$.

4. ELEMENTS OF FIRST EXCESS LEVEL THEORY

Consider again the arrival compound process (τ^a, A) and the point process $\tau^v = \{\tau_k^v; k = 0,1,2,\cdots\}$. Suppose that both τ^a and τ^v are independent of each other and start from the same origin $0 = \tau_0^a = \tau_0^v$. We now construct the following compound process based on the arrival marked process (τ^a, A). Let X^v be the sequence of increments X_k^v of (τ^a, A) taken over the subsequent points τ_k^v, $k = 1,2,\ldots$. In Section 3 it was shown that the

elements of X^ν are independent and identically distributed discrete-valued random variables with common PGF $\mathbf{E}[u^{X_k^\nu}] = \varphi^\nu(u)$. Thus, $A^\nu = \{A_k^\nu; k = 1,2,\ldots\}$, where

$$A_k^\nu = \sum_{j=1}^k X_j^\nu,$$

is a renewal process whose inciements form position independent marks for the marked point process (τ^ν, A^ν). For future needs we define A_0^ν to be a discrete-valued random variable independent of A^ν having, in general, a different distribution, given by $\mathbf{E}[u^{A_0^\nu}] = \varphi_0^\nu(u)$. We extend (τ^ν, A^ν) by augmenting the mark A_0^ν to the existing sequence of marks $\{A_k^\nu\}$ and assigning it to the point $\tau_0^\nu = 0$. We will use the same (τ^ν, A^ν) to denote this *delayed* compound renewal process. In practice, $A_0^\nu = Q_0^\nu$ will be the initial value of the queue (in our case strictly less than r) when the server goes on vacation.

The following terminology and notation is essentially due to [17]. The vacation sequence ends at time τ_Δ^ν, when the cumulative queue length hits r. This instant is called the *first passage time* of (τ^ν, A^ν). The queue level at τ_Δ^ν, previously denoted by Q_Δ^ν, is called the *first excess level of* (τ^ν, A^ν). In our case, it is convenient to extend the notion of the "first excess level" as follows. We set the first excess level to be equal to Q_1^{bs} consistent with definition (3.1). This will allow us to use a more compact notation. The random variable Δ is said to be the *termination index* of (τ^ν, A^ν). Formally, $\Delta = \Delta(r) = inf\{k: A_k^\nu \geq r\}$. In our applications, we will be interested in the queue length, Q_1^{bs} defined in (3.1), at the beginning of the first service, which, as noted, coincides with the first excess level or initial queue length.

We seek an analytic representation of the functional

$$E = \mathbf{E}[\xi^\Delta e^{-\theta\tau_\Delta^\nu} u^{Q_1^{bs}}] \tag{4.1}$$

to which we apply the transformation

$$D_p(\cdot)(x) = (1 - x) \sum_{p\geq 0} x^p(\cdot), \, |x| < 1. \tag{4.2}$$

Denote

$$\mathscr{E} = \mathscr{E}(\xi,\theta,u,; x) = D_r(E)(x). \tag{4.3}$$

The functional E can be retrieved from $\mathscr{E}$ by the application of the operator

$$\mathfrak{D}_x^m(\cdot) = lim_{x\to 0} \frac{1}{m!} \frac{\partial^m}{\partial x^m} \frac{1}{1-x}(\cdot), \, m = 0,1,\cdots, \tag{4.4}$$

$$\mathfrak{D}_x^m(\cdot) = 0, \text{ for } m < 0, \tag{4.5}$$

to $\mathscr{E}$. (Note that, in the above notation, both p and x are dummy and are used to emphasize the variables to which they are applied.)

For various special cases throughout the remainder of this paper we state a few elementary properties of the operator $\mathcal{D}^m$:

*P*1) $\mathcal{D}_y^m$ is a linear operator with fixed points at every constant function.

P2) For any function g, analytic at zero,

$$\mathcal{D}_y^m(y^k g(y)) = \begin{cases} 0, & m < k \\ \mathcal{D}_y^{m-k}(g(y)), & m \geq k \end{cases}$$

The following statement, due to Dshalalow [17], is expressed in terms of the current notation and is adapted to our system.

Proposition 1 The functional $\mathcal{E}$ satisfies the formula

$$\mathcal{E} = \varphi_0^v(u) - \varphi_0^v(ux)\frac{1 - \xi h(\theta)\varphi^v(u)}{1 - \xi h(\theta)\varphi^v(ux)}, \tag{4.6}$$

which yields

$$\mathbf{E}[\xi^\Delta e^{-\theta\tau_\Delta^v} u^{Q_1^{bs}}]$$

$$= \varphi_0^v(u) - (1 - \xi h(\theta)\varphi^v(u))\mathcal{D}_x^{r-1}\left\{\frac{\varphi_0^v(ux)}{1 - \xi h(\theta)\varphi^v(ux)}\right\}, \tag{4.7}$$

where $h(\theta) = \mathbf{E}[e^{-\theta\tau_1^v}]$ is the Laplace-Stieltjes transform of the first vacation segment. One special case applies directly to our system. We will consider $\mathbf{E}^i[u^{Q_1^{bs}}]$. In this case, $A_0^v = Q_0^v = i$ and consequently, $\varphi_0^v(u) = u^i$. By properties P1 and P2, formula (4.7) reduces to

$$\mathbf{E}[\xi^\Delta e^{-\theta\tau_\Delta^v} u^{Q_1^{bs}}]$$

$$= u^i - u^i(1 - \xi h(\theta)\varphi^v(u))\mathcal{D}_x^{r-i-1}\left\{\frac{1}{1 - \xi h(\theta)\varphi^v(ux)}\right\}. \tag{4.8}$$

Observe that the last expression is consistent with the functional's initial value when $Q_1^{bs} = Q_0^s = i$. In particular, the PGF of Q_1^{bs}, denoted $L_i(u)$, is given by

$$L_i(u) = \mathbf{E}^i[u^{Q_1^{bs}}] = u^i - u^i(1 - \varphi^v(u))\mathcal{D}_x^{r-i-1}\left\{\frac{1}{1 - \varphi^v(ux)}\right\}. \tag{4.9}$$

The conditional PGF of the termination index is

$$\mathbf{E}^i[\xi^\Delta] = 1 - (1 - \xi)\mathcal{D}_x^{r-i-1}\left\{\frac{1}{1 - \xi\varphi^v(x)}\right\}, \tag{4.10}$$

which yields

$$\Delta_i = \mathbf{E}^i[\Delta] = \mathcal{D}_x^{r-i-1}\left\{\frac{1}{1-\varphi^\nu(x)}\right\}. \tag{4.11}$$

From (4.7) it also follows that the Laplace-Stieltjes transform of the first passage time, τ_Δ^ν, is

$$\mathbf{E}^i[e^{-\theta\tau_\Delta^\nu}] = 1 - (1 - h(\theta))\Delta_i. \tag{4.12}$$

Therefore, the mean value of the first passage time is

$$\mathbf{E}^i[\tau_\Delta^\nu] = \mu^\nu\Delta_i. \tag{4.13}$$

Interestingly, this result appears to follow directly from the well-known Wold's equation; however, the use of Wold's equation is not applicable here, since Δ obviously depends on the length of each vacation segment.

From (4.9) and (4.13) we have that the conditional mean value of Q_1^{bs} is

$$L_i'(1) = \mathbf{E}^i[Q_1^{bs}] = i + \lambda\alpha\mathbf{E}^i[\tau_\Delta^\nu] = i + \rho^\nu\Delta_i \tag{4.14}$$

(where we denoted $\rho^\nu = \lambda\alpha\mu^\nu$), which gives the average number of arrivals per one vacation segment.

Next we introduce $Q_1^{incr} = Q_1^{bs} - Q_0^s$ and call it the *increment of the first excess level* above the initial value of the queue. Note that Q_1^{incr} is equal to zero if $Q_0^s \geq r$. From (4.14) it follows that the conditional mean value of the total increment,

$$\mathbf{E}^i[Q_1^{incr}] = \mathbf{E}^i[Q_1^{bs}] - i = \rho^\nu\Delta_i, \tag{4.15}$$

where $\lambda\alpha$ is the average number of arrivals per unit of time, and $\rho^\nu\Delta_i$ is the average number of arrivals per one full vacation. Again it's tempting to arrive at this result from Wold's equation but again it does not apply.

5. MORE PRELIMINARIES

Formula (3.3) precludes the necessity to treat special transformations of PGFs. Suppose that a nonnegative random variable X valued in $\{0,1,\ldots\}$ has its PGF $f(u) = \mathbf{E}[u^X]$. For a nonnegative integer p define the operator

$$\mathcal{H}^p f(u) = \mathbf{E}[u^{(X-p)^+}]. \tag{5.1}$$

Lemma 2 $\mathcal{H}^p$ is a linear operator with fixed points at constant functions and such that

$$\mathcal{H}^p f(u) = u^{-p}f(u) + \mathcal{D}_x^{p-1}\{f(x) - u^{-p}f(ux)\}. \tag{5.2}$$

Proof Let $P_X = \sum_{n=0}^{\infty} f_n \epsilon_n$ and denote

$$T_p(u) = \sum_{k=p}^{\infty} u^k f_k. \tag{5.3}$$

Then,

$$\mathcal{H}^p f(u) = 1 - T_p(1) + u^{-p} T_p(u). \tag{5.4}$$

Now we apply the transformation D_p defined in (4.2) to $T_p(u)$ and get

$$D_p T_p(u,x) = f(u) - xf(ux). \tag{5.5}$$

To extract T_p from (5.5), we apply operator $\mathcal{D}^p$ to $D_p T_p(u,x)$. Then, because of property P2 we have

$$T_p(u) = f(u) - \mathcal{D}_x^{p-1}\{f(ux)\}. \tag{5.6}$$

The latter, after substitution into (5.4), yields the lemma. □

We now return to the evaluation of $P_i(u)$, defined in Section 3. From (3.1–3.3) it follows that

$$P_i(u) = \mathbf{E}^i[u^{Q_1^s}] = \mathbf{E}^i[u^{Q_1^{res}}]\mathbf{E}^i[u^{X_1^v}],\ i = 0,1,\ldots.$$

Specifically, for $i \geq r$,

$$P_i(u) = \mathbf{E}^i[u^{Q_1^s}] = u^{(i-R)^+}\ \mathbf{E}^i[u^{X_1^v}].$$

By Lemma 2,

$$\mathbf{E}^i[u^{Q_1^{res}}] = \mathcal{H}^R L_i(u) = u^{-R} L_i(u) + \mathcal{D}_x^{R-1}\{L_i(x) - u^{-R} L_i(ux)\}, \tag{5.7}$$

From which it follows that

$$\mathbf{E}^i[Q_1^{res}] = i - R + \rho^v \Delta_i + \mathcal{D}_x^{R-1}\{RL_i(x) - xL_i'(x)\} \tag{5.8}$$

Thus, we have

$$P_i(u) = \varphi^v(u)\mathcal{H}^R L_i(u),\ i \geq 0. \tag{5.9}$$

and hence

$$\mathcal{H}^R L_i(u) = u^{(i-R)^+},\ i \geq r. \tag{5.10}$$

6. STATIONARY PROBABILITY DISTRIBUTION OF THE EMBEDDED PROCESS

From (3.6), (5.7), (5.9) and (5.10), it is easy to obtain a variant of the Pollaczek-Khintchine formula:

$$P(u) = \varphi^{\nu}(u)\frac{\sum_{i=0}^{R-1} p_i\{u^R\mathcal{H}^R L_i(u) - u^i\}}{u^R - \varphi^{\nu}(u)}. \tag{6.1}$$

By Rouché's theorem, the denominator, $D(u) = u^R - \varphi^{\nu}(u)$, of $P(u)$ has exactly R zeros in the closed unit disk, $\bar{B}(0,1)$ (including $u = 1$). Since $P(u)$ is analytic in $B(0,1)$ and continuous on its boundary $\partial B(0,1)$ (because $P(u)$ is absolutely convergent here) the numerator, $N(u) = \sum_{i=0}^{R-1} p_i\{u^R\mathcal{H}^R L_i(u) - u^i\}$, of $P(u)$ must also vanish at all those zeros. It is also known that roots on the boundary are simple. Let z_s be all zeros of $D(u)$ (ordered arbitrarily), with their respective multiplicities m_s, $s = 0,1,\ldots,S$, with $z_0 = 1$, $m_0 = 1$, and such that $\sum_{s=0}^{S} m_s = R - 1$. Then, for each s, $N(u)$ vanishes at z_s and so do its $m_s - 1$ derivatives. This gives us R linear equations in unknowns p_i, $i = 0,1,\ldots,R-1$: For $s = 1,\ldots,S$ we have that

$$\sum_{i=0}^{R-1} p_i \frac{d^j}{du^j}\{u^R\mathcal{H}^R L_i(u) - u^i\}\Big|_{u=z_s} = 0, j = 0,\ldots, m_s - 1, s = 1,\ldots,S \tag{6.2}$$

and for $z_0 = 1$,

$$\sum_{i=0}^{R-1} p_i\{R - i + \mathbf{E}^i[Q_1^{rest}]\} = R - \rho^s \tag{6.3}$$

or

$$\sum_{i=0}^{R-1} p_i(\rho^{\nu}\Delta_i + \mathcal{D}_x^{R-1}\{RL_i(x) - xL_i'(x)\}) = R - \rho^s, \tag{6.4}$$

The uniqueness of the solution $p_0,\ldots,p_{R-1}$ follows by arguments similar to those in [10].

The results of this section are summarized in the following theorem.

Theorem 3. In the $M^X/G^{(\tau,R)}/1$-type quorum vacation model with nonexhaustive service and multiple vacations the embedded queueing process selected over the instants of service completions is ergodic if and only if $\rho^s = \lambda\alpha\mu^s < R$. Under this condition, the PGF $P(u)$ satisfies formulas (6.1–6.4).

7. APPLICATIONS

Define $\rho_i = \lambda\alpha\mathbf{E}^i[\tau_1^s]$, $i \geq 0$. $\mathbf{E}^i[\tau_1^s]$ is the conditional mean value of the first service cycle which, for $i \geq r$, coincides with μ^s, the mean service time. For $i < r$, it is the sum of μ^s and the mean first passage time. Therefore, ρ_i gives us the expected value of arrivals during the first service cycle, given that the initial queue length was i. Let $\boldsymbol{\rho} = (\rho_i;\ i \geq 0)$. Then, the value $\boldsymbol{p\rho}$ gives the expected number of arrivals during a service cycle in the steady state. Following [10], we call $\boldsymbol{p\rho}$ the *system intensity*, which, in our case, equals

$$\boldsymbol{p\rho} = \rho^s + \rho^v \sum_{i=0}^{r-1} p_i\Delta_i. \tag{7.1}$$

Define

$$l = \lim_{n\to\infty} \mathbf{E}^i[min\{Q_n^{bs}, R\}] \tag{7.2}$$

where Q_n^{bs} is the queue length at the beginning of the nth service, and call it the *server load*. This is the mean batch of customers taken for service in the steady state. Using standard probability arguments and formulas (4.13), (4.14), (5.8), and (6.3) one can show that the system intensity and the server load coincide. (See, for example, Dshalalow [10].) This gives the so-called "conservation law of the system" in equilibrium.

References

1. D.P. Heyman, Optimal Operating Policies for M/G/1 Queueing Systems, *Oper. Res.*, **16**, (1968), 362–382.
2. M.L. Chaudhry, and J.G.C. Templeton, *A First Course in Bulk Queues*, John Wiley, New York, 1983.
3. B. Doshi, Single-Server Queues with Vacations. In: *Stochastic Analysis of Computer and Communication Systems*, (H. Takagi, ed.) Elsevier Science Publ. B.V. (North-Holland), 1990, pp. 217–265.
4. L. Abolnikov and J.H. Dshalalow, On a Multilevel Controlled Bulk Queueing System $M^X/G^{r,R}/1$, *Journ. Appl. Math. Stoch. Anal.*, **5**, (**3**) (1992), 237–260.
5. M.F. Neuts, A General Class of Bulk Queues with Poisson Input, *Ann. Math. Stat.* (1967), 759–770.
6. H.S. Lee, and M.M. Srinivasan, Control Policies for the $M^X/G/1$ Queueing System, *Manag. Sci.*, **35**, (**6**) (1989), 708–721.
7. H. W., Lee, S.S. Lee, and K.C. Chae, Operating Characteristics of $M^X/G/1$ Queue with N-Policy, *QUESTA*, **15** (1994), 387–399.
8. J. Loris-Teghem, On Vacation Models with Bulk Arrivals, *Belg. Journ. Oper. Res. Stat. Comp. Sci.*, **30**, (**1**) (1989), 53–66.
9. L. Abolnikov and J.H. Dshalalow, A First Passage Problem and Its Applications to the Analysis of a Class of Stochastic Models, *Journ. Appl. Math. Stoch. Analysis*, **5**, (**1**) (1992), 83–98.
10. J.H. Dshalalow, First Excess Level Analysis of Random Processes in a Class of Stochastic Servicing Systems with Global Control, *Stoch. Analysis and Appl.*, **12**, (**1**) (1994), 75–101.
11. D.C.R. Muh, A Bulk Queueing System under *N*-Policy with Bilevel Service Delay Discipline and Start-Up Time, *Journ. Appl. Math. Stoch. Analysis*, **6**, (**4**), (1993), 359–384.
12. D.C.R. Muh, On a Classes of *N*-Policy Multilevel Control Queueing Systems, Doctoral thesis, Florida Tech., Melbourne, 1994.
13. J.H. Dshalalow, On a First Passage Problem in General Queueing Systems with Multiple Vacations, *Journ. Appl. Math. Stoch. Analysis*, **5**, (**2**) (1992), 177–192.

14. M. Jacob and T. Madhusoodanan, Transient Solution for a Finite Capacity $M/G^{a,b}/1$ Queueing System with Vacations to the Server, *QUESTA*, **2** (1987), 381–386.
15. J.H. Dshalalow, On Termination Time Processes. In: *Studies in Applied Probability*, (J. Galambos and J. Gani, eds.) Essays in honor of Lajos Takács, *Journ. of Appl. Prob.*, **31A** (1994), 325–336.
16. J.H. Dshalalow, First Excess Levels of Vector Processes, *Journ. Appl. Math. Stoch. Anal.*, **7**, (**3**) (1994), 457–464.
17. J.H. Dshalalow, Excess Level Processes in Queueing. In: *Advances in Queueing: Theory, Methods, and Open Problems*. (J.H. Dshalalow, ed.) CRC Press, Boca Raton, 1995, pp. 233–252.
18. L. Abolnikov, J.H. Dshalalow, and A.M. Dukhovny, Stochastic Analysis of a Controlled Bulk Queueing System with Continuously Operating Server: Continuous Time Parameter Process, *Stat. Prob. Letters*, **16** (1993), 121–128.