ON A FIRST PASSAGE PROBLEM IN GENERAL QUEUEING SYSTEMS WITH MULTIPLE VACATIONS[1]

JEWGENI H. DSHALALOW

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

ABSTRACT

The author studies a generalized single-server queueing system with bulk arrivals and batch service, where 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 modulated by the system and the service is assumed to be state dependent. One of the essential part in the analysis of the system is the employment of new techniques related to the first excess level processes. A preliminary analysis of such processes and recent results of the author on modulated processes enabled the author to obtain all major characteristics for the queueing process explicitly. Various examples and applications are discussed.

Key words: Queueing process, modulated random measure, first excess level process, first passage time, semi-Markov process, embedded Markov chain, semi-Markov modulated marked Poisson process, equilibrium, semi-regenerative process.

AMS Subject Classification: 60K10, 60K15, 60K25.

1. INTRODUCTION

The class of queueing systems, where a server (or servers) periodically goes on "vacation" was offered in many works published in seventies because of its connection with congestion phenomena in local area networks. The interest in such systems has been further enhanced in the eighties and the beginning of the nineties due its applicability in communications, computer and production systems. Such models also apply to manufacturing processes that exhibit uninterruptible maintenance tasks, for instance, tool changes or alterations of a flexible manufacturing system. Vacation models have been studied in many queueing systems of types $M/G/1$, $GI/G/1$ and closed systems (see Dshalalow [8]).

Since the first publications on vacation models over twenty years ago the number of works on this topic has increased tremendously and as a result several survey papers appeared (see for instance Doshi [10]) classifying various systems and their modifications. One of the most frequently studied types of vacation models are those with so-called *exhaustive service* and *general nonexhaustive service*. The exhaustive service refers to the systems, where the server takes a vacation only when the queue of primary customers is empty. General nonexhaustive service refers to many other systems, where the server may go on vacation even if there are some customers waiting for service. For example, if the server takes batches of customers of some minimum size (say r), he prefers not to wait until the queue accumulates so many and it goes on vacation. Upon his return the server checks if the queue length is of an appropriate size, and if not, leaves the system again, and so on. Such systems are also classified as having *multiple vacations*. Clearly, this particular nonexhaustive system generalizes a class of exhaustive systems with $r = 1$. But there are also some other modifications of general nonexhaustive models, such as those with *Bernoulli schedules* (see Ramaswamy and Servi [16]). It is assumed in such a system that after each return to the system after a vacation the server may interrupt his vacation cycle regardless of how many customers are in the system, in the general case, with a probability dependent on the number of vacations taken (see Kella [13]). Many papers in the literature on queueing theory dealt with such models under different assumptions to the input stream, service discipline and the waiting room capacity. Queueing processes, busy period processes, and other processes were analyzed in those papers. The methods applied included decomposition (see Foorman and Cooper [11] and Shanthikumar [17]) and supplementary variables techniques (see Takagi [19]).

In the present paper the author considers a class of queueing systems of type $M^X/G^r/1$ with general nonexhaustive service and multiple vacations. It is assumed that if after completion of a service there are at least r customers in the waiting room the server takes a batch of customers of size r and begin their service. Otherwise, the server leaves the system for a vacation and upon its return the server checks if during its absence more customers arrived at the system and there are enough customers to pack the batch. If this is still not the case the server again leaves the system and so on. It is assumed that the input stream is compound Poisson. Under these assumptions, the model studied in this article generalizes a class of multiple vacations systems treated by various authors in the recent past (Abolnikov, Dshalalow and Dukhovny [3], Harris and Marchal [12], Lee and Srinivasan [14], Minch [15] and Shanthikumar [17]).

One of the central problems in the analysis of such systems is the behavior the queueing

process at the time the server begins processing groups of customers. Excluding trivial cases, it is clear that at those instants of time, the queueing process is more likely to exceed level r (*first excess level*) than just to reach it, and thus the results on "level crossing analysis" of processes [4,6,18] are inappropriate. Fortunately, as the author will show it, the paper [1] by Abolnikov and Dshalalow contains relevant results for the analysis of excess level processes needed for multiple vacations systems.

The author will also show that the analysis is well suited for generalized versions with modulated input and state dependent service times. In the present article queuing processes with continuous and discrete time parameters are treated. The author establishes necessary and sufficient conditions for the ergodicity of the processes, and by the use of the semi-regenerative analysis, the author derives a simple and explicit relation between the stationary distributions for the both processes. The embedded Markov process is analyzed by methods developed by Abolnikov and Dukhovny [2] and Abolnikov, Dshalalow and Dukhovny [3]. The invariant probability measure of the transition probability matrix of the embedded process is obtained in terms of the probability generating function in an explicit form. Various examples and applications are discussed.

2. MODULATED INPUT PROCESS

2.1 Definition (Dshalalow [9]). Let E be a locally compact and σ-compact space with a countable base. Let $\mathcal{M}$ be the space of all Radon measures on the Borel σ-algebra $\mathfrak{B}(E)$, and let $\mathfrak{M}$ be the Borel σ-algebra in $\mathcal{M}$ generated by the vague topology $\mathfrak{T}_\nu$. Denote by $\mathfrak{C}_K$ the space of all continuous functions on E with compact support and denote by $\mathfrak{M}$ the Baire σ-algebra in $\mathcal{M}$ generated by all maps $\mu \mapsto \int f d\mu$, $f \in \mathfrak{C}_K$.

(*i*) Let $\{\Omega, \mathfrak{F}, P, \xi(t),\ t \in E\} \to \Psi$ be a stochastic process on E and let ξ_ω denote the ω-section of ξ. Then for $\Gamma \subseteq \Psi$ and $B \in \mathfrak{B}(E)$ we define $\mathfrak{Y}_\Gamma = B \cap \xi_\omega^{-1}(\Gamma)$ and call it the *holding time of ξ in Γ on set B*. For each fixed ω, $\mathfrak{Y}_\Gamma$ is a measurable subset of E which can be measured by any Radon measure on $\mathfrak{B}(E)$. In general, $\mathfrak{Y}_\Gamma$ is a mapping from Ω into $\mathfrak{B}(E)$ which can be made a *random set* with respect to the σ-algebra $\{\Lambda \subseteq \mathfrak{B}(E): \mathfrak{Y}_\Gamma^{-1}(\Lambda) \in \mathfrak{F}\}$.

(*ii*) Consider for each j a marked random measure $Z_j = \sum_i S_{ij} \varepsilon_{\tau_{ij}}$ (where ε_x denotes the Dirac point mass) with mark space $\{0,1,...\}$ and introduce

(2.1a)
$$Z^\xi = Z^\xi(\omega, B) = \sum_{j=0}^{\infty} Z_j(\mathfrak{Y}_{\{j\}}).$$

The random measure $Z^\xi: (\Omega, \mathfrak{F}) \to (\mathcal{M}, \mathfrak{M})$ is called a *marked random measure modulated by the*

process ξ. The marked random measure Z^ξ can be more vividly represented in the form

$$Z^\xi = \sum_{n=1}^{\infty} S_{n\xi} \varepsilon_{\tau_{n\xi}}.$$

(*iii*) Consider the following special case. Let Z_j be a compound random measure (with mark space $\{0,1,\dots\}$)

(2.1b)
$$Z_j = \sum_i S_{ij} \varepsilon_{\tau_{ij}},$$

obtained from the underlying counting measure $N_j = \sum_i \varepsilon_{\tau_{ij}}$ by independent marking, i.e. for each $\{j\}$, $\{X_{ij} = S_{ij} - S_{i-1,j}; i = 1,2,\dots\}$ is a sequence of independent and identically distributed random variables with common mean denoted α_j. Assume that N_j is a Poisson counting measure with mean measure $\mu_j = \lambda_j \mathcal{L} \in \mathcal{M}$, where $\mathcal{L}$ denotes the Lebesgue measure and λ_j is a positive constant. Then, we call such a marked random measure modulated by ξ a *compound Poisson random measure modulated by the process* ξ. $\square$

Assume that $E = \mathbb{R}_+$ with its usual topology. Let $\{\Omega, \mathcal{F}, (P^x)_{x \in \Psi}, Q(t); \ t \in E\} \to \Psi = \{0,1,\dots\}$ be a stochastic process describing the evolution of the number of units in a single-server queueing system, and let $\{T_n; n \in \mathbb{N}_0, T_0 = 0\}$ be the sequence of successive completions of service and $Q_n = Q(T_n + 0)$.

2.2 Notation. Denote $C = \sum_{n=0}^{\infty} \varepsilon_{T_n}$ and $\xi(t) = Q(T_{C([0,t])} + 0)$, $t \in E$. Then the input is a compound Poisson process *modulated by* ξ according to definition 2.1 (*iii*), from which it follows that customers arrive at random instants of time $\tau_{n\xi}$, $n = 1,2,\dots$, (with arrival intensity $\lambda_{\xi(t)}$), that form a point process modulated by ξ with $X_{n\xi(t)}$ as the ith batch size of the input flow. Thus, in our case $X = \{X_{n\xi(t)}\}$ is an integer-valued doubly stochastic sequence describing the sizes of groups of entering units. We assume that, given $\xi(t)$, all terms of X are independent and identically distributed. Denote $a_{\xi(t)}(z) = E[z^{X_{n\xi(t)}}]$, $n = 1,2,\dots$, the generating function of the nth component of the process X.

3. CONTROL OF THE SERVICING PROCESS

At time $T_n + 0$ the server can carry a group of units of size r and it takes that many for service if available. Otherwise (that is if $Q_n < r$), the server leaves for vacation for a random period of time with probability distribution function $V_{Q_n} \in \{V_0, V_1, \dots, V_{r-1}\}$, after which the server returns to the system and leaves again if the total number of customers in the system is still less than r. The server will immediately start processing a group of customers if upon its return the queue has accumulated at least r customers. [Since the server leaves the system for a random time

and it does not interrupt its vacation even if the queue length has accumulated the desired level, and because customers arrive in batches, it is more likely in general that the queue level by that time exceeds the level r rather than it will equal r.] The service time of that group of customers is supposed to last a random time σ_{n+1} distributed $B_{Q_n} \in \mathfrak{B} = \{B_0, B_1, ...\}$, where $\mathfrak{B}$ is a given set of arbitrary probability distribution functions with finite means $\{b_0, b_1, ...\}$. By P^i or E^i we denote the conditional expectation induced by the initial measure ε_i. We denote

$$B_i^*(\theta) = E^i[e^{-\theta\sigma_1}], \; \Re(\theta) \ge 0.$$

The capacity of the waiting room is assumed to be unlimited.

We need more details and formalism of how exactly vacations and service are functioning. Suppose that the server leaves the system at time T_n and returns at time $T_n + \mathcal{V}_1^n$ reaching the system with $Q(T_n + \mathcal{V}_1^n) < r$. Then the server leaves the system again and it is coming back and forth until upon its arrival at $T_n + \mathcal{V}_1^n + ... + \mathcal{V}_\nu^n$ the queue length for the first time reaches or exceeds level r, where ν is an integer-valued random variable counting the total number of server vacations prior to the service begin. [Of course, $\nu(n) = \nu$ also depends upon n.] Let $Y_{n1}, Y_{n2}, ...$ be the increments of the arrival process over time intervals $(T_n, T_n + \mathcal{V}_1^n]$, $(T_n, T_n + \mathcal{V}_2^n], ...$. Then,

$$S_{nk} = Y_{n0} + Y_{n1} + ... + Y_{nk}, \text{ where } Y_{n0} = Q_n,$$

is a modulated integer-valued delayed renewal process with the increments $Y_{n1}, Y_{n2}, ...$, distributed in accordance with a common generating function $q_{Q_n}(z)$ which can easily be computed as

(3.1) $$q_{Q_n}(z) = V_{Q_n}^*(\lambda_{Q_n} - \lambda_{Q_n} a_{Q_n}(z)),$$

where $V_i^*(\theta)$ denotes the Laplace-Stieltjes transform the distribution function V_i. Thus $\nu(n) = inf\{k: S_{nk} \ge r\}$ and the instant $\mathcal{T}_n = T_n + \mathcal{V}_1^n + ... + \mathcal{V}_\nu^n$ is the *first passage time* (after T_n) of the queueing process to reach or to exceed level r. In a situation when $Q_n \ge r$, we may denote $\mathcal{T}_n$ to be T_n to extend the notion of the first passage time for all cases.

Finally, figuring that $Z^\xi(\sigma_n)$ gives the total number of customers that arrive during the period of nth service, σ_n, we obtain the following relation for process $\{Q_n\}$:

(3.2) $$Q_{n+1} = \begin{cases} S_{n\nu} - r + Z^\xi(\sigma_{n+1}), & Q_n < r \\ Q_n - r + Z^\xi(\sigma_{n+1}), & Q_n \ge r. \end{cases}$$

4. FIRST EXCESS LEVEL PROCESSES

Throughout sections 4-8, we will use some basic results on a first passage problem stated and developed in Abolnikov and Dshalalow [2].

Let $\{\Omega, \mathcal{F}, P\}$ be a probability space. Consider on this probability space a delayed renewal process $\{S_n = \mathcal{V}_0 + \mathcal{V}_1 + ... + \mathcal{V}_n ; n \geq 0\}$ valued in a set $\Psi \preceq \mathbb{N}$. Let $\mathcal{V}_0 = S_0 = i$ a.s. (for some $i \in \mathbb{N}_0$) and let $\mathcal{V}_1$ be distributed in accordance with the generating function $q(z) = E[z^{\mathcal{V}_1}]$, analytic inside the open unit ball $B(0,1)$ centered at the origin and continuous on its boundary $\partial B(0,1)$, and with finite mean $q = E[\mathcal{V}_1]$.

For a fixed integer $s \geq 1$ we will be interested in the behavior of the process $\{S_n\}$ and some related processes about level s.

The following terminology from [2] will be used throughout the paper.

4.1 Definitions.

(*i*) Denote $\nu = inf\{k \geq 0 : S_k \geq r\}$ and call it the *index of the the first excess* of level r.

(*ii*) The random variable S_ν is called the *level of the first excess* of r.

(*iii*) The random variable $\mathcal{T}$ is known as the *first passage time* of $\{S_k\}$ of level r.

Let

(4.1a)
$$\gamma_s^{(i)}(\theta, z) = E[e^{-\theta \mathcal{T}} z^\nu], \quad \mathfrak{g}_s^{(i)}(\theta, z) = E[e^{-\theta \mathcal{T}} z^{S_\nu}], \quad G_{is}(\theta, z)$$
$$= \sum_{j \geq 0} E[e^{-\theta \mathcal{V}_j} z^{S_j} I_{U_{s-1}} \circ S_j], \quad e(\theta) = E[e^{-\theta \mathcal{T}}] = \frac{\lambda}{\lambda + \theta}.$$

where $U_p = \{0, 1, ..., p\}$ and I_A is the indicator function of a set A. We call $G_r(\theta, z)$ the *generator of the first excess level*. We will also use the following functionals of marginal processes:

(4.1b) $\gamma_s^{(i)}(z) = \gamma_s^{(i)}(0, z),$

(4.1c) $\mathfrak{g}_s^{(i)}(z) = \mathfrak{g}_s^{(i)}(0, z),$

(4.1d) $G_{is}(z) = G_{is}(0, z).$

It is readily seen that $G_{is}(z)$ is a polynomial of $(r - 1)$th degree.

We formulate the main theorems from Abolnikov and Dshalalow [2] and give formulas for the joint distributions of the first passage time and the random variables listed in 4.1 (*i-iii*). To adopt these results for the processes to be treated in the next sections we additionally assume that the parameters (as λ_i, $a_i(z)$, α_i and $q_i(z)$) of the point, compound and vacation processes introduced sections 2 and 3 may depend on index i (in agreement with the modulation). Consequently, these parameters will be indexed by i. The random variable $\mathcal{T}$ will then become the first term $\mathcal{T}_1$ of the point process $\{\mathcal{T}_n\}$ of first passage times.

4.2 Theorem. *The functional* $\gamma_r^{(i)}(\theta, z)$ *(of the first passage time* $\mathcal{T}_1$ *and of the index* ν_1 *of*

the first excess level) satisfies the following formula:

(4.2a)
$$\gamma_r^{(i)}(\theta, z) = \begin{cases} e_i(\theta) z \mathfrak{D}_x^{r-i-1}\left\{\dfrac{1 - q_i(x)}{(1-x)(1 - z e_i(\theta) q_i(x))}\right\}, & i < r \\[12pt] 1, & i \geq r. \end{cases}$$

where

(4.2b)
$$\mathfrak{D}_x^k = \lim_{x \to 0} \frac{1}{k!} \frac{\partial}{\partial x}.$$

From formula (4.2a) we immediately obtain that the conditional mean value of the first excess index equals

(4.3)
$$\bar{\gamma}_r^{(i)} = \begin{cases} \mathfrak{D}_x^{r-i-1}\left\{\dfrac{1}{(1-x)[1 - q_i(x)]}\right\}, & i < r \\[12pt] 0, & i \geq r. \end{cases}$$

Specifically, the conditional mean value of the first passage time is

(4.4)
$$E^i[\mathcal{T}_1] = \frac{1}{\lambda_i} \bar{\gamma}_r^{(i)}.$$

4.5 Theorem. *The generator $G_{ir}(\theta, z)$ of the first excess level can be determined from the following formula:*

(4.5a)
$$G_{ir}(\theta, z) = \begin{cases} z^i \mathfrak{D}_x^{r-i-1}\left\{\dfrac{1}{(1-x)[1 - e_i(\theta) q_i(xz)]}\right\}, & i < r \\[12pt] 0, & i \geq r. \end{cases}$$

The use of the term "generator of the first excess level" is due to the following theorem.

4.6 Theorem. *The functional $\mathfrak{g}_r^{(i)}(\theta, z)$ (of the first passage time and of the first excess level) can be determined from the formula*

(4.6a)
$$\mathfrak{g}_r^{(i)}(\theta, z) = z^i - [1 - e_i(\theta) q_i(z)] G_{ir}(\theta, z).$$

4.7 Remark. To obtain the functionals of the marginal processes defined in (4.1b-4.1d) we set $e_i(\theta) = 1$ in formulas (4.2a), (4.5a) and (4.6a).

5. SEMI-REGENERATIVE ANALYSIS

In this section we will analyze the queueing process $\{Q(t)\}$. It will be shown that this process is semi-regenerative relative to the point process $\{T_n\}$ and we will obtain its ergodicity conditions.

We need the following notion introduced in Dshalalow [7].

5.1 Definition. Let T be a stopping time for a stochastic process $\{\Omega, \mathfrak{F}, (P^x)_{x \in \Psi}, Q(t); t \geq 0\}$ $\to (\Psi, \mathfrak{B}(\Psi))$. $\{Q(t)\}$ is said to have the *locally strong Markov property* at T if for each bounded random variable $\zeta: \Omega \to \Psi^r$ and for each Baire function $f: \Psi^r \to \mathbb{R}$, $r = 1, 2, ...$, it holds true that

$$E^x[f \circ \zeta \circ \theta_T \,|\, \mathfrak{F}_T] = E^{Z_T}[f \circ \zeta] \quad P^x\text{-}a.s. \text{ on } \{T < \infty\},$$

where θ_y is the shift operator.

From the nature of the input process and relation (3.2) it follows that $\{\Omega, \mathfrak{F}, (P^x)_{x \in \Psi}, Q(t); t \geq 0\} \to \Psi = \{0, 1, ...\}$ possesses a *locally strong Markov property* at T_n, where T_n is a stopping time relative to the past of the process $\sigma(Q(y); y \leq t)$, for each $n = 1, 2, ...$. Thus the embedded process Q_n is a homogeneous Markov chain with the transition probability matrix denoted $A = (a_{ij}; i, j \in \Psi)$. In the next section we will show that Q_n is irreducible and aperiodic, and that under a certain (necessary and sufficient) condition, it is recurrent-positive. We assume that this condition is met and denote by P the invariant probability measure of the operator A. Consequently, the two-dimensional process $\{Q_n, T_n\}$ is a Markov renewal process. Therefore, we conclude that the process $\{\Omega, \mathfrak{F}, (P^x)_{x \in \Psi}, \xi(t); t \geq 0\} \to \Psi$, defined in notation 2.2, is the minimal semi-Markov process associated with Markov renewal process $\{Q_n, T_n\}$ and therefore, following definition 2.1, the input process $\{\Omega, \mathfrak{F}, (P^x)_{x \in \Psi}, Z^\xi([0,t]); t \geq 0\} \to \Psi$ is a compound Poisson process modulated by the semi-Markov process $\{\xi(t)\}$. Then the value $\beta_i = E^i[T_1]$ is obviously the mean sojourn time of ξ in state $\{i\}$. Denote $\beta = (\beta_x; x \in \Psi)^T$. In one of the next sections we show that the value $P\beta$ (scalar product of the invariant probability measure and the vector of the mean sojourn times of the process ξ) which is called the *mean inter-renewal time* of the Markov renewal process, is finite.

5.2 Notation. Let

(5.2a) $$\delta_{xs}(t) = P^x\{Z^\xi([0, t]) = s \,|\, T_1 > t\} .$$

Then, given that $\xi(0) = x$ and because Z^ξ is not modulated by a new value of ξ, the input process takes on value Z_x (introduced in definition 2.1 (ii)). Therefore, we have

(5.2b) $$\delta_{xs}(t) = P\{Z_x([0, t]) = s\} . \qquad\qquad \square$$

Let $\{\Omega, \mathfrak{F}, (P^x)_{x \in \Psi}, Q(t); t \geq 0\} \to (\Psi, \mathfrak{B}(\Psi))$ be a semi-regenerative process relative to the sequence $\{T_n\}$ of stopping times. Introduce the probability

$$K_{ik}(t) = P^i\{Q(t) = k, T_1 > t\}, i, k \in \Psi.$$

We will call the functional matrix $K(t) = (K_{ik}(t); i, k \in \Psi)$ the *semi-regenerative kernel*. The following proposition holds true. [We set the value of any sum is zero whenever the lower index is greater

than the upper index.]

5.3 Proposition. *The semi-regenerative kernel satisfies the following equations:*

(5.3a) $K_{jk}(t) =$

$$\begin{cases} \delta_{j,k-j}(t), & 0 \leq j \leq k < r \\ \sum_{s=r-j}^{k-j} \int_0^t \varphi_j(s+j, t-u) \delta_{j,k-j-s}(u)(1 - B_j(u)) \, du, & 0 \leq j < r \leq k \\ \delta_{j,k-j}(t)[1 - B_j(t)], & r \leq j \leq k \\ 0, & 0 \leq k < j, \ j > 0, \end{cases}$$

where $\delta_{j,k}$ is as defined in (5.2a) or (5.2b) and φ_j denotes the joint probability density function of the first excess level S_ν and the first passage time $\mathfrak{T}_1$.

Proof. The above assertion follows from direct probability arguments.   □

Now we are ready to apply the Main Convergence Theorem.

5.4 Theorem. *Let $\{\Omega, \mathfrak{F}, (P^x)_{x \in \Psi}, Q(t); \ t \geq 0\} \to (\Psi, \mathfrak{B}(\Psi))$ be a semi-regenerative process relative to $\{T_n\}$ and let $K(t)$ be the semi-regenerative kernel. Suppose that the associated Markov renewal process is ergodic, the embedded Markov chain Q_n is ergodic (and its invariant probability measure is P) and that $K(t)$ is integrable over $\mathbb{R}_+$. Then the stationary distribution $\pi = (\pi_x; \ x \in \Psi)$ of $Q(t)$ exists and it can be determined in terms of its generating function $\pi(z)$ as follows. Denote $H = (h_{jk}; \ j,k \in \Psi) = \int_0^\infty K(t) dt$, $h_j(z)$ the generating function of jth row of matrix H and $h(z) = (h_j: j \in \Psi)^T$. Then,*

(5.4a)
$$\pi(z) = \frac{Ph(z)}{P\beta},$$

where P is the stationary probability vector of Q_n and β is the vector of mean sojourn times of the semi-Markov process ξ.

One of the main results of this section follows.

5.5 Theorem. *Given an equilibrium condition for the embedded process Q_n, the stationary distribution $\pi = (\pi_x; x \in \Psi)$ of the queueing process $Q(t)$ exists; it is independent of any initial distribution and is expressed in terms of the generating function $\pi(z)$ of π as:*

$$P\beta\pi(z) = \sum_{j \geq 0} p_j h_j(z), \ with$$

(5.6b)
$$h_j(z) = \frac{1}{\lambda_j} K_j(z) G_{jr}(z) + z^j \Delta_j(z), \ 0 \leq j < r,$$

(5.6c)
$$h_j(z) = z^j \Delta_j(z), \ r \leq j, \ and$$

(5.6d)
$$\Delta_j(z) = \frac{1 - K_j(z)}{\lambda_j(1 - a_j(z))},$$

where $G_{jr}(z)$ is the generator of the first excess level.

Proof. By the reasoning similar to those in Dshalalow [7] the Markov renewal process (Q_n, T_n) is ergodic. By theorem 5.4 the semi-regenerative process $Q(t)$ has a unique stationary distribution π. From proposition 5.3 we can see that the semi-regenerative kernel $K(t)$ is integrable over $\mathbb{R}_+$. Following 5.4, we evaluate the integrated semi-regenerative kernel H and then vector $h(z)$. First we find that

(5.6e) $$\sum_{p=j}^{\infty} z^p \int_0^{\infty} \delta_{j,p-j}(u)[1 - B_j(u)]\, du = z^j \Delta_j(z)\ .$$

Then it follows that

(5.6f) $$h_j(z) = z^j \frac{1}{\lambda_j} \mathfrak{D}_x^{r-j-1} \left\{ \frac{1}{(1-x)(1-q_j(xz))} \right\} + \Delta_j(z) G_{jr}^{(r)}(z),\ 0 \le j < r\ ,$$

where $G_{jr}^{(r)}(z)$ denotes the tail of the generating function $G_{jr}(z)$ summing its terms from r to ∞. However, by is easy to show that $G_{jr}^{(r)}(z)$ and $G_{jr}(z)$ coincide. Then it appears that

(6.6g) $$h_j(z) = z^j \Delta_j(z),\ j \ge r.$$

Then the statement of the theorem follows.   □

6. INVARIANT PROBABILITY MEASURE OF OPERATOR A

Let $A_i(z)$ denotes the generating function of ith row of matrix A. Since $A_i(z) = E^i[z^{Q}1]$ we obtain from (3.2) that

(6.1) $$A_i(z) = K_i(z) z^{-r} \mathfrak{g}_r^{(i)}(z),\ i \in \Psi,\ \qquad \text{where}$$

(6.1a) $$K_i(z) = B_i^*(\lambda_i - \lambda_i a_i(z)),$$

$B_i^*(\theta)$, $Re(\theta) \ge 0$, is the Laplace-Stieltjes transform of the probability distribution function B_i, and $\mathfrak{g}_r^{(i)}$ satisfies formulas (4.5a) and (4.6a), taking into account remark 4.7.

For analytical convenience and without considerable loss of generality we can drop the modulation of the input process and service control when the queue length exceeds a fixed (perhaps very large) level N. In other words, we assume that

(AS) $B_i(x) = B(x), \beta_i(\theta) = \beta(\theta), K_i(z) = K(z), b_i = b, \lambda(i) = \lambda_i = \lambda, a_i(z) = a(z), a_i = a,\ i > N$,

$N \ge r - 1$, where $\alpha_i = a_i'(1)$, $i \in \Psi$.

Given assumption (AS), it can be shown that the transition probability matrix A is reduced to a form of the $\Delta_{r,N}$-matrix introduced and studied in [2]. According to Abolnikov and Dukhovny [2], the Markov chain $\{Q_n\}$ is recurrent-positive if and only if

(6.2) $$\frac{d}{dz} A_i(z)\Big|_{z=1} < \infty,\ i = 0,1,...,N,$$

and

(6.2a)
$$\frac{d}{dz}\, K(z)\Big|_{z\,=\,1} < r\,.$$

Condition (6.2) is obviously met and condition (6.2a) is equivalent to

(6.2b)
$$\rho = \lambda a b < r\,.$$

Therefore, given that $\rho < r$, the Markov chain $\{Q_n\}$ is ergodic. Let $P = (p_x\,;x \in \Psi)$ be the invariant probability measure of operator A and let $P(z)$ be the generating function of the components of vector P. Denote $\bar{B}(0,1) = \{z \in \mathbb{C}\colon \|z\| \le 1\}$. Now we formulate the main result of this section.

6.3 Theorem. *The embedded queueing process* $\{Q_n\}$ *is ergodic if and only if* $\rho < r$. *Under this condition,* $P(z)$ *is determined by the following formula:*

(6.3a)
$$P(z) = \frac{\sum_{i\,=\,0}^{N} p_i \Big\{ K_i(z)\mathfrak{g}_r^{(i)}(z) - z^i K(z) \Big\}}{z^r - K(z)}\,,$$

Probabilities $p_0,...,p_N$ *form the unique solution of the following system of linear equations:*

(6.3b)
$$\sum_{i\,=\,0}^{N} p_i \frac{d^k}{dz^k}\{z^{-r} K_i(z)\mathfrak{g}_r^{(i)}(z) - z^i\}\Big|_{z\,=\,z_s} = 0,\ \ k = 0,...,k_s - 1\,,\ \ s = 1,...,S,$$

(6.3c)
$$\sum_{i\,=\,0}^{N} p_i[\rho_i - \rho - i + \bar{\mathfrak{g}}_r^{(i)}] = r - \rho\,,$$

where z_s *are the roots of* $z^{N+1} - z^{N+1-r} K(z)$ *in the region* $\bar{B}(0,1)\backslash\{1\}$ *with their multiplicities* k_s *such that* $\sum_{s\,=\,1}^{S} k_s = N$, *and where*

(6.3d)
$$\rho_i = \lambda_i \alpha_i b_i\,,$$

and $\bar{\mathfrak{g}}_r^{(i)}$ *can be determined from (4.5a) and (4.6a).*

Proof. Formula (6.3a) follows from $P(z) = \sum_{i \in \Psi} p_i A_i(z)$ and (6.1-6.1a). It is easy to modify formula (6.3a) into

(6.3e)
$$\sum_{i\,=\,N+1}^{\infty} p_i z^{i-N-1} = \frac{\sum_{i\,=\,0}^{N} p_i \{K_i(z) z^{-r} \mathfrak{g}_r^{(i)}(z) - z^i\}}{z^{N+1} - z^{N+1-r} K(z)}\,,$$

so that the function in the left-hand side of (6.3f) is analytic in $B(0,1) = \{z \in \mathbb{C}\colon \|z\| < 1\}$ and continuous on its boundary $\partial B(0,1)$. According to [2], for $\rho < r$, the function $z \mapsto z^r - K(z)$ has exactly r zeros in $\bar{B}(0,1)$ (counting with their multiplicities); all zeros located on the boundary ∂B (including 1), are simple. Therefore, the denominator in the right-hand side of (6.3e) must have exactly N roots in the region $\bar{B}(0,1)\backslash\{1\}$. This fact along with $(P,1) = 1$ (which yields (6.3c)) leads to equations (6.3b-6.3d).

The uniqueness of $p = \{p_0,...,p_N\}$ follows from the same considerations as in [7]. $\qquad\square$

7. OTHER PARAMETERS OF THE SYSTEM

7.1 Definitions.

(i) Let $\beta_j = E^j[T_1]$ (the mean sojourn time of the process $\{\xi(t)\}$ in state $\{j\}$) and $\beta = (\beta_j; j \in \Psi)^T$. Then we will call the value $P\beta$ the *mean service cycle of the system*, where P denotes the stationary probability distribution vector of the embedded queueing process $\{Q_n\}$.

(ii) Let $\alpha = (\alpha_x; x \in \Psi)^T$, $\lambda = (\lambda_x; x \in \Psi)^T$ and let $\rho = \alpha*\beta*\lambda$ be the Hadamard (entry-wise) product of vectors α, β and λ. We call the scalar product $P\rho$ the *stationary intensity of the system*.

Observe that the notion "intensity of the system" goes back to the classical $M/G/1$ system, when $P\rho$ reduces to $\rho = \lambda b$. It is noteworthy that in some systems both capacities coincide.

7.2 Proposition. *Given the equilibrium condition $\rho < r$, the mean service cycle can be determined from the following expression:*

$$(7.2a) \qquad P\beta = b + \sum_{j=0}^{N} P_j(b_j - b + \tfrac{1}{\lambda_j}\bar{\gamma}_r^{(i)})$$

Proof. Obviously, $\beta_j = b_j + \bar{\gamma}_r^{(i)}/\lambda_j$. The statement follows after elementary algebraic transformations.   □

From (7.3a) we similarly derive the stationary intensity of the system.

$$(7.3) \qquad P\rho = \rho + \sum_{j=0}^{N} P_j(\rho_j - \rho + \alpha_i\bar{\gamma}_r^{(i)}).$$

7.4 Lemma. *The expected number of units $\bar{g}_s^{(i)} - i$ that arrive during an idle period of the server in equilibrium satisfies the following expression:*

$$(7.4a) \qquad \bar{g}_s^{(i)} - i = \alpha_i \bar{\gamma}_s^{(i)}, \; s = 1,...,r, \; i \in \Psi.$$

Proof. Formula (7.4a) directly follows from (4.5a) and (4.6a). [Intuitively, it seems that the same result could follow from the Wald's formula applied directly to $E^i[S_{\nu_0}] - i$. However, this would be unjustified, since the values of random variables contained in the sum S_{ν_0} are not independent of ν_0, as some counterexamples show it.]   □

7.5 Theorem. *Given the equilibrium condition $\rho < r$, the intensity of the system $P\rho$ and server capacity coincide and equal r.*

Proof. According to the description of our model the server capacity is r. Now, the statement of the theorem follows from definition 7.1 (ii), equation (6.3c) and lemma 7.4.   □

7.6 Remark. The equation (6.3c) can now be substituted by a more elegant relation

(7.6a)
$$P\rho = r.$$

8. EXAMPLES AND SPECIAL CASES

8.1 Theorem. *Given the assumption* (AS), *the probability generating function of the stationary probability vector π of the process $Q(t)$ can be determined from the following formula:*

$$P\beta\pi(z) = \Delta(z)P(z) + \sum_{j=0}^{N} p_j(h_j(z) - z^j\Delta(z)), \text{ with}$$

$$h_j(z) = \frac{1}{\lambda_j} K_j(z) G_{jr}(z) + z^j\Delta_j(z), \ 0 \leq j < r,$$

$$h_j(z) = z^j\Delta_j(z), \ r \leq j \leq N, \text{ and}$$

$$\Delta_j(z) = \frac{1 - K_j(z)}{\lambda_j(1 - a_j(z))},$$

where $P(z)$ is the generating function of P, $P\beta$ is determined in proposition 7.2, $G_{jr}(z)$ is the generator of the first excess level, and $\Delta(z)$ is defined as $\Delta_j(z)$ with all subscripts dropped.

8.2 Examples.

(*i*) By dropping the condition of modulation of the input we have from theorem 7.5 that

(8.2a)
$$P\beta = \frac{r}{\lambda\alpha}.$$

Observe that the same result holds true by retaining a "vague modulation", i.e. assuming that $\lambda_i = \lambda$ and $\alpha_i = \alpha$ but having no further restriction to the generating functions $a_i(z)$.

(*ii*) Assuming that the input is an orderly modulated Poisson process, in other words if $a_j(z) = z$, but retaining all other assumptions we arrive at the result

$$h_j(z)\frac{1}{\lambda_j}(1 - z)z^{-j} = 1 - \hat{K}_j(z),$$

where $\hat{K}_j(z) = z^{(j-r)^+} K_j(z)$ (with $f^+ = sup\{f, 0\}$ for a function f); and then we obtain from (8.2a) that

(8.2b) $P\beta\pi(z)(1 - z) = \frac{1}{\lambda}[1 - K(z)]P(z) + \sum_{j=0}^{N} p_j z^j\left(\frac{1}{\lambda_j} - \frac{1}{\lambda}\right) - \sum_{j=0}^{N} p_j\left(\frac{1}{\lambda_j}z^r\hat{K}_j(z) - \frac{1}{\lambda}z^j K(z)\right),$

$$P\beta = b + \sum_{j=0}^{N} p_j[(b_j - b) + \frac{1}{\lambda_i}(r - i)^+].$$

By dropping the modulation of the input process we obtain from (8.2a) that $P\beta = \frac{r}{\lambda}$ and from (8.2b) that

(8.2c)
$$\pi(z) = \frac{1 - z^r}{r(1 - z)} P(z).$$

(*iii*) By dropping the modulation but retaining bulks of the input, service control and state dependent service delay, we obtain from theorem 8.1 and (8.2a) by means of routine calculus more general result than that in (8.2c):

(8.2d)
$$\pi(z) = \frac{\alpha(1 - z^r)P(z)}{r[1 - a(z)]}.$$

(*iv*) By using obvious probability arguments we can derive the probability density function of an idle period in the steady state:

$$u \longmapsto \frac{\sum_{i=0}^{r-1} p_i \gamma^{(i)}(\theta, 1)}{\sum_{i=0}^{r-1} p_i},$$

The mean idle period $\mathfrak{I}$ in the steady state is then

(8.2e)
$$\mathfrak{I} = \frac{\sum_{i=0}^{r-1} p_i \frac{1}{\lambda_i} \bar{\gamma}_r^{(i)}}{\sum_{i=0}^{r-1} p_i}.$$

(*v*) Formula (8.2e) and theorem 8.1 allow to derive the mean busy period $\mathfrak{B}$ in the equilibrium. Clearly $\sum_{i=0}^{r-1} \pi_i$ is the probability that the server idles. On the other hand, it also equals $\frac{\mathfrak{I}}{\mathfrak{I} + \mathfrak{B}}$. Thus we have

$$\mathfrak{B} = \frac{\mathfrak{I} \sum_{n=r}^{\infty} \pi_n}{\sum_{i=0}^{r-1} \pi_i}.$$

(*vi*) If the input is a stationary compound Poisson process (i.e. nonmodulated) then its intensity is $\alpha\lambda$, which is also the mean number of arriving units per unit time. In the case of a modulated input process its intensity is no longer a trivial fact. We define the intensity of any random measure Z by the formula $\kappa = \lim_{t \to \infty} \frac{1}{t} \mu_t(x)$, where $\mu_t(x) = E^x[Z([0,t])]$. We will apply the formula from Dshalalow [9] for more general Poisson process modulated by a semi-Markov process:

$$\kappa = \frac{P\rho}{P\beta},$$

where by theorem 7.5 $P\beta = r$ and $P\beta$ satisfies (7.2a). Thus we have that:

(8.22f)
$$\kappa = \frac{r}{P\beta}.$$

A trivial special case appears when we assume the vague modulation of the input defined in (*i*) and therefore use formula (8.2a) combining it with formula (8.2f). Then $\kappa = \lambda\alpha$. Specifically $\kappa = \lambda\alpha$ for the input without modulation, as it should be.

REFERENCES

[1]. Abolnikov, L. and Dshalalow, J., A first passage problem and its applications to the analysis of a class of stochastic models, *Journ. Appl. Math. Stoch. Anal.*, 5, No. 1, 83-98, 1992.

[2]. Abolnikov, L. and Dukhovny A., Markov chains with transition delta-matrix: ergodicity conditions, invariant probability measures and applications, *Journ. Appl. Math. Stoch. Anal.*, 4, No. 4, 335-355, 1991.

[3]. Abolnikov, L., Dshalalow, J. and Dukhovny, A., On stochastic processes in a multilevel

control bulk queueing system, *Stoch. Analysis and Appl.*, **10**, No. 2, 155-179, 1992.

[4]. Arnold, B. and Groenevald, R., On excess life in certain renewal processes, *Journ. Appl. Probab.*, **18**, 379-389, 1981.

[5]. Çinlar, E., *Introduction to Stochastic Processes*, Prentice Hall, Englewood Cliffs, N.J., 1975.

[6]. Cohen, J., On up- and downcrossings, *Journ. Appl. Probab.*, **14**, 405-410, 1977.

[7]. Dshalalow, J., Single-server queues with controlled bulk service, random accumulation level, and modulated input, to appear in *Stoch. Analysis and Applications*.

[8]. Dshalalow, J., On single-server closed queues with priorities and state dependent parameters, *Queueing Systems*, **8**, 237-254, 1991.

[9]. Dshalalow, J., On modulated random measures. *Journ. Appl. Math. Stoch. Anal.*, **4**, No.4, 305-312, 1991.

[10]. Doshi, B., Single server queues with vacations, in *Stochastic Analysis of Computer and Communication Systems*, edited by H. Takagi, Elsevier North Holland, 217-265, 1990.

[11]. Foorman, S. and Cooper, R., Stochastic Decomposition in an $M/G/1$ queues with generalized vacations, *Oper. Res.*, **33**, 1117-1129, 1985.

[12]. Harris, C. and Marchal, W., State dependence in $M/G/1$ server-vacation models, *Oper. Research*, **36**, No. 4, 560-565, 1988.

[13]. Kella, O., Optimal control of the vacation scheme in an $M/G/1$ queue, *Oper. Res.*, **38**, No. 4, 724-728, 1990.

[14]. Lee, H. and Srinivasan, M., Control policies for the $M^X/G/1$ queueing system, *Manag. Sci.*, **35**, No. 6, 708-721, 1989.

[15]. Minch, D., Transient solution for some exhaustive $M/G/1$ queues with generalized independent vacations, *European Journ. Oper. Res.*, **36**, 197-201, 1988.

[16]. Ramaswamy, R. and Servi, L, The busy period of the $M/G/1$ vacation model with Bernoulli schedule, *Communic. Stat. Stochast.. Models*, **4**, 507-521 (1988).

[17]. Shanthikumar, J., On stochastic decomposition in $M/G/1$ type queues with generalized server vacations, *Oper. Res.*, **36**, No. 4, 566-569, 1988.

[18]. Shanthikumar, J.G., Level crossing analysis of priority queues and a conservation identity fro vacation models, *Nav. Res. Loqist.*, **36**, 797-806, 1989.

[19]. Takagi, H., Time-dependent analysis of $M/G/1$ vacation models with exhaustive service, *Queueing Systems*, **6**, 369-390, 1990.