A SINGLE-SERVER QUEUE WITH RANDOM ACCUMULATION LEVEL[1]

JEWGENI H. DSHALALOW
Department of Applied Mathematics, Florida Institute of Technology
Melbourne, FL 32901, USA

ABSTRACT

The author studies the queueing process in a single-server bulk queueing system. Upon completion of a previous service, the server can take a group of random size from customers that are available. Or, the server can wait until the queue attains a desired level.

The author establishes an ergodicity criterion for both the queueing process with continuous time parameter and the imbedded process. Under this criterion, the author obtains explicit formulas for the stationary distributions of both processes by using semi-regenerative techniques.

Key words: Random Group Size, Single-Server Queue, Queueing Process, Imbedded Markov Chain, Semi-Regenerative Process.

AMS Subject Classification: Primary 60K10, 60K25, Secondary 90B22, 90B25.

1. INTRODUCTION

In a large class of bulk queueing models, the server takes groups of a fixed size for service if enough group members are available; otherwise, it waits until the queue reaches a desired (fixed) level. Several versions of such systems are considered in Dshalalow/Russel [4] and Dshalalow/Tadj [5]. We call such systems *queues with fixed accumulation level.* Practically more attractive and versatile, but analytically more complicated, is a system with a *random accumulation level.* In such a system, the server capacity is a random number generated by the completion of previous service and this number is the desired group size to be taken for service. The server will therefore rest until the queue accumulates that many customers if that group size is unavailable by the time the server becomes free.

---

[1]Received: March 1991 Revised: June 1991

For instance, for shipment of certain goods not only are transportation units of different capacity used, but arriving units can also be partially occupied. Units can take some of the load and move that quantity farther, or wait until the load reaches a specified level. Although such situations are most common in air and surface transportation, postal delivery, inventory-transportation systems and assembly lines, there are other real systems of the same nature that can be modeled by queues with random accumulation levels. For example, a computer user needs a specific task to be performed on several parallel or networked computers or processors. The job can only be started when all necessary computer components become free. So in this case the job to be done will be regarded as a server and computers will play the role of the customers. Again each particular job needs a different (random) number of computers. Thus the situation can be described in terms of a model with a random accumulation level.

In the present paper, the author introduces a class of stochastic models with a random accumulation level and studies the queueing process with discrete and continuous time parameter. In both cases the author establishes the ergodicity criterion and derives explicit formulas for the limiting distribution of the processes.

2. DESCRIPTION OF THE SYSTEM AND NOTATION

Let $Q(t)$ denote the number of customers in a single-server queueing system at time $t \geq 0$ and let $Q_n = Q(t_n + 0)$, $n = 1, 2, \ldots$, where t_n is the moment of time when the server completes the processing of the nth group of customers. At time $t_n + 0$ the server can carry a group of customers of size c_{n+1} and it takes that many for service if available. If not available, that is if $Q_n < c_{n+1}$, the server prefers to rest as long as necessary for the queue to accumulate to the level of c_{n+1}. Only then does it begin to process a group of the appropriate size, with the pure service time lasting σ_{n+1}. We assume that each of the sequences $\{c_n\}$ and $\{\sigma_n\}$ are families of independent identically distributed random variables, independent of each other and of the input stream.

The probability distribution of c_1 is given by $g_k = P\{c_1 = k\}$, $k = 1, \ldots, r$. The random variable σ_1 has an arbitrary probability distribution function B, with $B(x) = 0$ for $x < 0$, and with a finite mean b. We denote

$$g(x) = E[x^{c_1}]$$
$$\beta(\theta) = \int_0^\infty e^{-\theta x} B(dx), \quad \Re(\theta) \geq 0.$$

The input stream is formed by an orderly stationary Poisson point process $\{\tau_n\}$ with intensity λ; and the capacity of the waiting room is assumed to be unlimited.

3. IMBEDDED PROCESS

Let $N(\cdot)$ denote the counting measure associated with the point process $\{\tau_n\}$. Denote $v_n = N(\sigma_n)$. Then the terms of the sequence $\{Q_n\}$ satisfy the following recursive relation:

$$(3.1) \qquad Q_{n+1} = \begin{cases} Q_n + (c_{n+1} - Q_n) + v_{n+1} - c_{n+1}, & Q_n < c_{n+1} \\ Q_n - c_{n+1} + v_n, & Q_n \geq c_{n+1}. \end{cases}$$

Clearly the process $\{\Omega, \mathfrak{F}, (P^x)_{x \in E}, Q(t); t \geq 0\} \to E = \{0,1,...\}$ possesses a *locally strong Markov property* at t_n (see Definition A.1 in Appendix), where t_n is a stopping time relative to the canonic filtering $\sigma(Q(y); y \leq t)$, $n = 1,2,...$. Thus the imbedded process $\{Q_n\}$ is a homogeneous Markov chain with the transition probability matrix $A = (p_{ij}; i,j \in E)$. Due to (3.1) the upper block $(p_{ij}; i = 0,1,...,r-1, j \in E)$ of A consists of purely positive elements, and the lower block of A is an upper triangular matrix (with all positive entries on the main diagonal and above it and all zero elements below the main diagonal). Clearly the Markov chain $\{Q_n\}$ is irreducible and aperiodic. According to Abolnikov and Dukhovny [2], A is a $\Delta_{r,r}$-matrix and the ergodicity of $\{Q_n\}$ is given by the following criterion.

3.1 Lemma (Abolnikov/Dukhovny [2]). *Let* $\{Q_n\}$ *be an irreducible aperiodic Markov chain with the transition probability matrix* A *in the form of a* $\Delta_{r,r}$-*matrix* (3.1). $\{Q_n\}$ *is recurrent-positive if and only if*

$$(3.2) \qquad \lim_{z \to 1: z \in B(0,1)} \frac{d}{dz} A_i(z) < \infty, \quad i = 0,1,...,r-1,$$

and

$$(3.3) \qquad \lim_{z \to 1: z \in B(0,1)} \frac{d}{dz} A_r(z) < r,$$

where $A_i(z)$ *is the generating function of ith row of the transition probability matrix* A *and* $B(z_0, \rho)$ *denotes an open ball in* $\mathbb{C}$ *centered at* z_0 *with radius* ρ.

3.2 Proposition. *The generating function* $A_i(z)$ *of ith row of the transition probability matrix* A *satisfies the following formula:*

$$(3.4) \qquad A_i(z) = \beta(\lambda - \lambda z) z^i \, G_i(\tfrac{1}{z}), \quad i \in E,$$

where

$$(3.5) \qquad G_i(z) = \begin{cases} \sum_{s=1}^{i} g_s z^s + \sum_{s=i+1}^{r} g_s z^i, & i < r \\ g(z) = \sum_{s=1}^{r} g_s z^s, & i \geq r. \end{cases}$$

Proof. Formulas (3.4) and (3.5) follow from (3.1) by use of standard probability calculus. $\square$

Now we turn to Lemma 3.1. While condition (3.2) is obviously satisfied, formula (3.3) applied to (3.4) and (3.5) leads to the following.

3.3 Theorem. *The imbedded Markov chain* $\{Q_n\}$ *is irreducible and aperiodic. It is recurrent positive if and only if*

(3.6) $\rho < \bar{g}$,

where $\rho = \lambda b$ *and* $\bar{g} = E[c_1]$ *is the mean server capacity.*

4. INVARIANT PROBABILITY MEASURE

Given the equilibrium condition (3.6), the invariant probability measure $P = (p_i; i \in E)$ of the operator A exists and equals the stationary distribution of the Markov chain $\{Q_n\}$. The following statement obviously holds true.

4.1 Lemma. *Let* $P(z)$ *denote the generating function of the invariant probability measure* P *of a transition probability matrix* A *of a homogeneous Markov chain* $\{Q_n\}$, *and let* $A_i(z)$ *denote the generating function of* ith *row of* A. *Then*

(4.1) $P(z) = \sum_{i \in E} A_i(z) p_i$.

Using lemma 4.1 and the ideas of the last two sections, we obtain the following main result.

4.2 Theorem. *Given the ergodicity condition in theorem 3.3, the generating function* $P(z)$ *of the stationary distribution of the imbedded queueing process* $\{Q_n\}$ *satisfies the following formula:*

(4.2) $$P(z) = \frac{\beta(\lambda - \lambda z) \sum_{i=0}^{r-1} p_i z^i \left[G_i(\tfrac{1}{z}) - g(\tfrac{1}{z}) \right]}{1 - g(\tfrac{1}{z}) \beta(\lambda - \lambda z)} ,$$

where G_i *is defined in* (3.5).

Although formula (4.2) contains r unknown probabilities, $p_0, ..., p_{r-1}$, they can be determined from an additional condition which yields relatively simple equations. The latter can be solved numerically.

4.3 Theorem. *The probabilities* $p_0, ..., p_{r-1}$ *satisfy the following system of linear equations:*

(4.3) $\sum_{i=0}^{r-1} p_i \; \dfrac{d^k}{dz^k} \; z^i [\, G_i(\tfrac{1}{z}) \beta(\lambda - \lambda z) - 1)] \Big|_{z \,=\, z_s} = 0$, $k = 0, ..., k_s - 1$, $s = 1, ..., S$

(4.4) $\sum_{i=0}^{r-1} p_i \sum_{s=i+1}^{r} g_s(s - i) = \bar{g} - \rho$

where G_i *satisfies formula* (3.5) *and* $\{z_s; s = 1, ..., S\}$ *is the set of roots of the function* $z^r - \beta(\lambda - \lambda z) \sum_{k=1}^{r} g_k z^{r-k}$ *inside the unit ball* $B(0,1)$ *with their multiplicities* k_s, *such that* $\sum_{s=1}^{S} k_s = r - 1$. *The system of equations* (4.3)-(4.4) *has a unique solution,* $p_0, ..., p_{r-1}$.

Proof. Formula (4.2) can be rewritten in the form

$$\sum_{i=r}^{\infty} p_i z^{i-r} = \frac{\sum_{i=0}^{r-1} p_i z^i \left(G_i(\tfrac{1}{z}) \beta(\lambda - \lambda z) - 1 \right)}{z^r - \sum_{k=1}^{r} g_k z^{r-k} \beta(\lambda - \lambda z)} .$$

The rest of the proof is similar to that of theorem 4.2 [1]. $\qquad\qquad\square$

4.4 Definition and Notation.

Let $\beta_i = E^i[t_1]$. This gives the expected length of the service cycle given that the initial queue length was equal to i. Let $\beta = (\beta_i;\ i \in E)^{\mathrm{T}}$. Then the scalar product $P\beta$ gives the value of the mean service cycle of the system in the stationary mode. We wish to call the ratio of the mean service cycle $P\beta$ and the mean inter-arrival time the *capacity of the system*. Thus the capacity of the system is defined as $\lambda P\beta$.

Earlier we denoted the mean server capacity by $\bar{g}$. Observe that for the classical $M/G/1$ queue the capacity of the system is $\lambda b + p_0 = 1$, which coincides with server capacity. Below we show we have this remarkable property in our case also, when the system is in the equilibrium.

4.5 Proposition. *Given the equilibrium condition, the capacity of the system $\lambda P\beta$ and server capacity $\bar{g}$ are equal.*

Proof. Evaluating β_i we have

$$\beta_i = b + I_{\{0,\,...,\,r-1\}}(i) \sum_{s=i+1}^{r} \tfrac{1}{\lambda} (s-i) g_s ,$$

where I_D is the indicator function of a set D. The statement follows from the last equation and formula (4.4). $\qquad\qquad\square$

5. ANALYSIS OF THE CONTINUOUS TIME PARAMETER PROCESS

From the discussion in section 3 and from definition A.1, it follows that $\{\Omega, \mathfrak{F}, (P^x)_{x\in E},$ $Q(t);\ t \geq 0\} \rightarrow (E, \mathfrak{B}(E))$ is a semi-regenerative process with conditional regenerations at points $t_n,\ n = 0,1,...,\ t_0 = 0.$ $\{\Omega, \mathfrak{F}, (P^x)_{x\in E},\ (Q_n, t_n):\ n = 0,1,...\} \rightarrow (E \times \mathbb{R}_+,\ \mathfrak{B}(E \times \mathbb{R}_+))$ is the associated Markov renewal process. Let $\mathfrak{F}(t)$ denote the corresponding semi-Markov kernel. With a very mild restriction to the probability distribution function B, we can have that the elements of $\mathfrak{F}(t)$ are not step functions and thus we can have (Q_n, t_n) aperiodic. By proposition 4.5 the mean inter-renewal time $P\beta$ of the Markov renewal process equals $\bar{g}/\lambda$ ($< \infty$). Therefore, following definition A.3, the Markov renewal process is ergodic given the condition $\rho < \bar{g}$.

Let $K(t)$ be the semi-regenerative kernel (see definition A.4). The following proposition holds true. [We will agree throughout the paper that the value of any sum is zero whenever the lower index is greater than the upper index.]

5.1 Proposition. *The semi-regenerative kernel satisfies the following formulas:*

$$(5.1) \quad K_{jk}(t) = \begin{cases} \sum_{s=j+1}^{min(k,r)} K_{jk}^{(s)}(t)g_s + \pi_{\lambda t}(k-j)[1-B(t)] \sum_{s=1}^{min(j,r)} g_s + \pi_{\lambda t}(k-j) \sum_{s=k+1}^{r} g_s, & 0 \leq j \leq k \\ 0 & , 0 \leq k < j, \end{cases}$$

where

$$(5.2) \quad K_{jk}^{(s)}(t) = \int_0^t e_{\lambda, s-j}(t-u)\pi_{\lambda u}(k-s)[1-B(u)]du, \quad 0 \leq j \leq s-1, \; 1 \leq s \leq min(k,r),$$

while $(\pi_u; u \in \mathbb{R}_+)$ *denotes the Poisson semi-group and* $e_{\lambda,k}$ *is a k-Erlang probability density function with parameter* λ.

Proof. The statement follows from probability arguments.   □

Now we are ready to apply the Main Convergence Theorem to the semi-regenerative kernel in the form of corollary A.6.

5.2 Theorem. *Given the equilibrium condition* $\rho < \bar{g}$ *for the imbedded process* $\{Q_n\}$, *the stationary distribution* $\pi = (\pi_x; x \in E)$ *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 *in the following formula:*

$$(5.3) \quad \pi(z)\bar{g}(1-z) = P(z)[1 - \beta(\lambda - \lambda z)] + \sum_{i=0}^{r-1} [G_i(z) - g(z)]p_i,$$

where $P(z)$ *is the generating function of* P *and* G_i *is defined in* (3.5).

Proof. Recall that the Markov renewal process (Q_n, t_n) is ergodic if $\rho < \bar{g}$. By corollary A.6 the semi-regenerative process $\{Q(t)\}$ has a unique stationary distribution π provided $\rho < \bar{g}$. From (5.1) and (5.2) we can see that the semi-regenerative kernel is Riemann integrable over $\mathbb{R}_+$. Thus following corollary A.6 we need to find the integrated semi-regenerative kernel H (which is done with routine calculus) and then generating functions $h_j(z)$ of all rows of H. We have

$$(5.4) \quad \lambda(1-z)h_j(z) = [1 - \beta(\lambda - \lambda z)]\left\{z^j \sum_{s=1}^{j} g_s + \sum_{s=j+1}^{r} g_s z^s\right\}$$
$$+ \sum_{s=j+1}^{r} g_s(z^j - z^s), \; 0 \leq j < r$$

$$(5.5) \quad \lambda(1-z)h_j(z) = z^j[1 - \beta(\lambda - \lambda z)], \; r \leq j.$$

Formula (5.3) now follows from proposition 4.5, formula (A.6) and expressions (5.4) and (5.5). □

5.3 Examples.

(*i*) It follows from (5.3) that

$$\pi_0 = \frac{p_0[1 - \beta(\lambda)]}{\bar{g}}.$$

(*ii*) Let $\mathfrak{J}$ denote the expected length of an idle period in the equilibrium. Then by direct probability arguments we have

$$\mathfrak{J} = \frac{\frac{1}{\lambda}\sum_{i=0}^{r-1} p_i \sum_{s=i+1}^{r} g_s(s-i)}{\sum_{i=0}^{r-1} p_i \sum_{s=i+1}^{r} g_s}.$$

By formula (4.4) the expression for $\mathfrak{I}$ is simplified to

$$(5.6) \qquad \mathfrak{I} = \frac{\bar{g} - \rho}{\lambda \sum_{i=0}^{r-1} p_i \sum_{s=i+1}^{r} g_s} .$$

(*iii*) Similarly we obtain the probability density function of an idle period in the equilibrium:

$$f(t) = \frac{\sum_{i=0}^{r-1} p_i \sum_{s=i+1}^{r} e_{\lambda,\,s-i}(t) g_s}{\sum_{i=0}^{r-1} p_i \sum_{s=i+1}^{r} g_s} ,$$

where $e_{\lambda,\,k}$ is a k-Erlang density function.

(*iv*) Let $\mathfrak{B}$ denote the expected value of a busy period in the equilibrium. Then the probability of reaching the server idle in the steady state can be given by the following expression:

$$\tilde{p} = \sum_{i=0}^{r-1} \pi_i \sum_{s=i+1}^{r} g_s = \frac{\mathfrak{I}}{\mathfrak{I} + \mathfrak{B}} .$$

Then, from the last equation and (5.6) we get a formula for the expected length of a busy period in the steady state:

$$(5.7) \qquad \mathfrak{B} = \frac{\mathfrak{I}(1 - \tilde{p})}{\tilde{p}} .$$

APPENDIX

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

A.2 Definition. A stochastic process $\{\Omega,\mathfrak{F},(P^x)_{x\epsilon E},\ Z(t);\ t \geq 0\} \to (E,\ \mathfrak{B}(E))$ with $E \preceq \mathbb{N}$ is called *semi-regenerative* if

a) there is a point process $\{t_n\}$ on $\mathbb{R}_+$ such that $t_n \to \infty$ $(n \to \infty)$ and such that each t_n is a stopping time relative to the canonic filtering $\sigma(Z_y; y \leq t)$,

b) the process $(Z(t))$ has the locally strong Markov property at t_n, $n = 1,2,...$,

c) $\{Z(t_n + 0), t_n;\ n = 0,1,...\}$ is a Markov renewal process.

A.3 Definition. Let (X_n, t_n) be an irreducible aperiodic Markov renewal process with a discrete state space E. Denote $\beta_x = E^x[t_1]$ as the mean sojourn time of the Markov renewal process in state $\{x\}$ and let $\beta = (\beta_x; x \in E)^\mathsf{T}$. Suppose that the imbedded Markov chain (X_n) is ergodic and that P is its stationary distribution. We call $P\beta$ the *mean inter-renewal time*. We call the Markov renewal process *recurrent-positive* if its mean inter-renewal time is finite. An irreducible aperiodic and recurrent-positive Markov renewal process is called *ergodic*.

A.4 Definition. Let $\{\Omega, \mathfrak{F}, (P^x)_{x \in E}, Z(t); t \geq 0\} \rightarrow (E, \mathfrak{B}(E))$ be a semi-regenerative process relative to the sequence $\{t_n\}$ of stopping times. Introduce the probability

$$K_{jk}(t) = P^j\{Z(t) = k, t_1 > t\}, \; j,k \in E.$$

We will call the functional matrix $K(t) = (K_{jk}(t); \; j,k \epsilon E)$ the *semi-regenerative kernel.*

A.5 Theorem (The Main Convergence Theorem, *cf.* Çinlar [3], p. 347). *Let* $\{\Omega, \mathfrak{F}, (P^x)_{x \in E}, Z(t); t \geq 0\} \rightarrow (E, \mathfrak{B}(E))$ *be a semi-regenerative stochastic process relative to the sequence* $\{t_n\}$ *of stopping times and let* $K(t)$ *be the corresponding semi-regenerative kernel. Suppose that the associated Markov renewal process is ergodic and that the semi-regenerative kernel is Riemann integrable over* $\mathbb{R}_+$. *Then the stationary distribution* $\pi = (\pi_x; x \in E)$ *of the process* $(Z(t))$ *exists and it is determined from the formula:*

$$(A.5) \qquad \qquad \pi_k = \frac{1}{P\beta} \sum_{j \in E} p_j \int_0^\infty K_{jk}(t) dt, \; k \in E.$$

A.6 Corollary. *Denote* $H = (h_{jk}; j,k \in E) = \int_0^\infty K(t) dt$ *as the integrated semi-regenerative kernel,* $h_j(z)$ *the generating function of jth row of matrix* H *and* $\pi(z)$ *as the generating function of vector* π. *Then the following formula holds true.*

$$(A.6) \qquad \qquad \pi(z) = \frac{1}{P\beta} \sum_{j \in E} p_j h_j(z)$$

Proof. From (A.5) we get an equivalent formula in matrix form, $\pi = \frac{PH}{P\beta}$. Finally, formula (A.6) is the result of elementary algebraic transformations. $\square$

REFERENCES

[1]. Abolnikov, L., Dshalalow, J. and Dukhovny, A., On some queue length controlled stochastic processes, *Journ. Appl. Math. Stoch. Analysis*, 3, No. 4, 227-244, 1990.

[2]. Abolnikov, L.M. and Dukhovny A.M., Necessary and sufficient conditions for the ergodicity of Markov chains with transition $\Delta_{m,n}$ ($\Delta'_{m,n}$) matrix, *Journ. Appl. Math. Simul.*, 1, No. 1, 13-24, 1987.

[3]. Çinlar, E., *Introduction to Stochastic Processes*, Prentice Hall, 1975.

[4]. Dshalalow, J. and Russel, G., On $M/G^Y/1$ type queue with fixed accumulation level, state dependent service and modulated input flow (in preparation).

[5]. Dshalalow, J. and Tadj, L., A queueing system with fixed accumulation level, random server capacity and capacity dependent service time, Techn. Rep. 01AM0691, Florida Institute of Technology, 1991.