Journal of Applied Mathematics and Stochastic Analysis
6, Number 4, Winter 1993, 359-384

A BULK QUEUEING SYSTEM UNDER N-POLICY WITH BILEVEL SERVICE DELAY DISCIPLINE AND START-UP TIME[1]

DAVID C. R. MUH[2]

Department of Operations Research
Florida Institute of Technology
Melbourne, FL 32901, USA

ABSTRACT

The author studies the queueing process in a single-server, bulk arrival and batch service queueing system with a compound Poisson input, bilevel service delay discipline, start-up time, and a fixed accumulation level with control operating policy. It is assumed that when the queue length falls below a predefined level r (≥ 1), the system, with server capacity R, immediately stops service until the queue length reaches or exceeds the second predefined accumulation level $N(\geq r)$. Two cases, with $N \leq R$ and $N \geq R$, are studied.

The author finds explicitly the probability generating function of the stationary distribution of the queueing process and gives numerical examples.

Key words: Embedded Queueing Process, N-policy, Bulk Arrival, Bilevel Service Delay Discipline, Start-Up Time, Transition Probability Matrix.

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

1. INTRODUCTION AND GENERAL DESCRIPTION OF MODELS

In modern computer communication networks, queueing theory is a useful tool to analyze node-to-node communication parameters. This is especially true in Packet Switched Computer Communication Systems. Nodes of many networks can be analyzed in terms of a standard M/G/1 queueing system. However, some situations require researchers to investigate complex M/G/1 queueing systems. Daigle [12] illustrates how the M/G/1 paradigm can be used to obtain fundamental insight into the behavior of a slotted-time queueing system that represents a statistical multiplexing system.

[1]Received: June, 1993. Revised: October, 1993.

[2]Postal address: 1767 Brookside St. N.E., Palm Bay, FL 32907.
E-Mail Address : *muh@zach.fit.edu*

In computer communication networks, it is common that the system stops servicing when the total number of messages in the input buffer falls below a preassigned level r, which is less than or equal to the server capacity R. The service is resumed if the system accumulates to at least r messages. This is known as bilevel (the level r and the server capacity R) service delay discipline or (r, R)-quorum. In some instances, the system waits until the total number of arriving messages becomes equal to or greater than another preassigned level N $(\geq r)$, so that upon reaching N, the system resumes servicing messages. This operating policy is known as N-(control operating) policy, and such system is noted as $M_N^x/G^{r,R}/1$.

In this paper, the author examines a single-server queueing system with a compound Poisson input stream and generally distributed service times, N-control operating policy, bilevel controlled service discipline, and start-up time. When total messages in the waiting queue equal or exceed the level r, the server may not be immediately available until some pre-service work warms up the system for service. As soon as the start-up time is over, the server starts the service of those messages in the waiting line.

We assume that the server capacity (R) is fixed, while r, the control level, and N, the control operating policy, can be adjusted to optimize system performance. Depending on the situation, N can be either selected from between levels r and R, or made greater than R. In the case of a very intense input, arriving units can be grouped within small intervals of time, thereby forming a bulk input.

Three different queueing models under N-policy are analyzed:

Model 1; with $r \leq N \leq R$ and its modification (Model 3 introduced at the end of this section).

Model 2; with $r \leq R \leq N$.

These models generalize two types of classical queueing systems: Model 1 establishes an additional control operating policy level N $(\geq r)$ such that when the queue length is either equal to or greater than N, the system changes from an idle to a busy state. This model generalizes the classical system with bilevel service delay discipline or (r, R)-quorum, M/G/1 bulk queueing and start-up time. Model 2, however, generalizes a system with single service delay discipline or R-quorum, M/G/1 bulk queueing and start-up time. Numerical examples for Model 1 and Model 2 are presented to demonstrate the application of obtaining an optimal solution for minimizing the system idle time. Results from both cases show that when server capacity is fixed, the optimal solution can be obtained if level r is at minimum and level N is at maximum.

The author also treats a modification of Model 1 which, for convenience, will be called "**Model 3.**" In this model, the server also accepts customers that arrive during the start-up time period in excess of the queue length (and under the bound R totaling the size of group taken for service). In many practical instances, this may be a more realistic scenario of Model 1.

2. RECENT RELATED WORK

In recent years, several papers have been published on the subject of M/G/1 models with N-policy. Heyman [14] studied an M/G/1 queue under "Control Operating Policy" in which the server stays idle when the queue length is empty and resumes its work when the queue length accumulates to a predefined level N (≥ 1). [Specifically, Heyman [14] showed that for the M/G/1 queue, there is a stationary optimal policy of the form: Turn the server on when N customers are present, and turn it off when the system is empty.] Bohm [6, 7] studied the transient mode of M/G/1 queue with N-policy. Baker [5] considered an M/M/1 queue under N-policy with exponentially distributed start-up time which the system requires before it changes the server state from idle to busy. Baker obtained the steady state distribution of the queue length. Borthakur, Medhi, and Gohain [8] studied M/M/1 queue under N-policy with general start-up time. Medhi and Templeton [16] studied an M/G/1 queue under N-policy with general start-up time.

An up-to-date extension of the M/G/1 standard system to the class of N-policy models would include the general start-up time and the N-policy itself. Perhaps the model studied in [16] was the most general in the available literature on M/G/1 systems under N-policy.

A few more systems [10, 15, 19] do not fall into this class of N-policy M/G/1 models but they are related either to N-policy or to the results obtained in the present paper. Chitkara and Kumar [10] studied N-policy for an Erlangian input system with reorientation period. These authors obtained the Laplace transforms of generating functions of the probabilities and the expressions for the steady state probabilities and the mean queue length in the system. Abolnikov and Dshalalow [2] studied an $M^x/G^{r,R}/1$ queueing system with a compound Poisson input modulated by a semi-Markov process, multilevel control service time and a queue length dependent service delay discipline. They found the stationary distribution of the queue process by using the results on excess level processes.

The present paper generalizes the existing class of all N-policy M/G/1 models to date (including [16] mentioned above). The author confirms Abolnikov/Dukhovny's [3] necessary and sufficient criterion of ergodicity of the embedded queueing processes and finds explicit formulas for its stationary distribution. A few numerical examples demonstrate the obtained results and discuss the best policies.

3. FORMALISM OF THE MODELS

Let $\{Q(t); t \geq 0\} \rightarrow \Psi = \{0,1,...\}$ be the total number of customers at time $t \geq 0$ in a single server queueing system with an infinite waiting room, and let $T_0 = 0,\ T_1,\ T_2,\ . . .,$ be the sequence of successive service completions of groups of customers. Defining $Q(t)$ as a right continuous process, we introduce the embedded process $Q_n = Q(T_n+) = Q(T_n)$, $n = 1,\ 2, ...$ Let the random variable σ_n be the service time of the nth group of customers.

Input Process.

Customers arrive at instants of time τ_n, $n = 1, 2, \ldots$, which form a Poisson point process with arrival intensity λ, in batches of sizes X_n, $n = 1, 2, \ldots$, as independent and identically distributed random variables with the common mean α, and the common probability generating function $a(z) = E\,[z^{X_n}]$, $n = 1, 2, \ldots$. Service times and sizes of groups to be served are independent of the queue length. Let $S_k = X_0 + X_1 + \ldots + X_k$ $(X_0 = Q_n)$. Denote $\nu_n = inf\{k \geq 0: S_k \geq N\}$. This is known [1] as the random index with which S_k first reaches or exceeds level N after the moment of time T_n at which the total number of customers in the system is Q_n. Note that τ_{ν_n} is the *first passage time* of the queue to reach or exceed N after T_n, and S_{ν_n} gives the total number of customers in the system at instant τ_{ν_n}.

Service Time and Service Discipline.

Let $r \le R$ be two integers, such that R is the server capacity and r is the minimal batch size the server is allowed to take. Let N be an integer such that, when queue length equals or exceeds N, the server changes from idle to busy state. N can either be placed in between r and R, or exceed R. At time $T_n + 0$ the server starts its $(n+1)$st service and carries a group of units of size $min\{Q_n, R\}$ if at least r customers are available. Otherwise, the server idles until the queue length reaches or exceeds level N for the first time. Obviously, if $Q_n \ge r$, $T_{n+1} - T_n$ coincides with the length of service time of the $(n+1)$st batch. In this case, we assume that the service lasts a random time σ_{n+1} with an arbitrary distribution function B and finite mean b. If $Q_n < r$, the server waits as long as necessary for the queue to accumulate to at least N units. The server activity resumes by the instant of time when the queue for the first time reaches or exceeds N. In this case, the system enters the start-up mode which lasts ξ_{n+1} (with an arbitrary distribution function D and finite mean d) followed by $(n+1)$st service. Given the queue length $Q_n < r$ and the server capacity R, a group of size $min\{S_{\nu_n}, R\}$ will "formally" be processed during the pure time σ_{n+1} of service after start-up time ξ_{n+1}. In this case, $T_{n+1} - T_n$ is the sum of server waiting time $\tau_{\nu_n} - T_n$, the actual service time σ_{n+1}, and the start-up time ξ_{n+1}. In Model 1, all newcomers during the start-up time are not accepted into the start-up servicing group. This is a somewhat artificial start-up service policy; however, it agrees with all known special models. In Model 2, newcomers entering the system during start-up time have no effect on the start-up servicing group, since S_{ν_n} is greater than or equal to R.

In Models 1 and 2, when the server begins processing the $(n+1)$st batch of units, its load can be defined as

$$L_{n+1}(Q_n) = \begin{cases} min\{S_{\nu_n}, R\}, & Q_n < r, \\ min\{Q_n, R\}, & Q_n \ge r. \end{cases} \tag{3.1}$$

A more realistic service policy can be employed as a modification of Model 1 by accepting new arrivals during the start-up time to the start-up servicing group, excluding those in excess of R. When the server begins processing the $(n+1)$st batch of units, its load can be defined as

(3.2) $$L_{n+1}(Q_n) = \begin{cases} min\{S_{\nu_n} + W_{n+1}, R\}, & Q_n < r, \\ min\{Q_n, R\}, & Q_n \geq r. \end{cases}$$

Denote $V_n = V(\sigma_n)$ the number of customers that arrive during service time σ_n and $W_n = W(\xi_n)$ the number of customers that arrive during the start-up time ξ_n. Then the values Q_n, $n = 1, 2, ...$, can be shown to satisfy the following recursive relations:

Model 1: $r \leq N \leq R$ (service does not include customers who arrived during a start-up time)

(3.3) $$Q_{n+1} = \begin{cases} (S_{\nu_n} - R)^+ + V_{n+1} + W_{n+1}, & Q_n < r, \\ (Q_n - R)^+ + V_{n+1}, & Q_n \geq r. \end{cases}$$

Model 2: $r \leq R \leq N$

(3.4) $$Q_{n+1} = \begin{cases} S_{\nu_n} - R + V_{n+1} + W_{n+1}, & Q_n < r, \\ (Q_n - R)^+ + V_{n+1}, & Q_n \geq r. \end{cases}$$

Model 3: $r \leq N \leq R$ (server may take some customers for service who arrived during a start-up time)

(3.5) $$Q_{n+1} = \begin{cases} (S_{\nu_n} + W_{n+1} - R)^+ + V_{n+1}, & Q_n < r, \\ (Q_n - R)^+ + V_{n+1}, & Q_n \geq r, \end{cases}$$

where $f^+ = sup\ \{f, 0\}$.

Note that all three models fall into the category of state dependent queueing systems. All of them have (r,R)-quorum and N-policy regarded as service discipline state dependency. The availability of the start-up time is a vague form of the general state dependent service time policy which, in its full power, was developed in [2]. Namely, it was assumed in [2] that random service times differ in their distributions, depending on the number of customers in the system. Inter-arrival times and sizes of arriving batches were governed by the queueing process at specified random times, so that it was a modification of the Poisson process, but with variable random intensities "modulated" by another process. "Input modulation" is the condition where inter-arrival times and sizes of arriving batches are also dependent on the queue length. Thus, the model studied in [2] had more flexible input and service time dependence than the models under our study. However, our models significantly generalize [2] in terms of more versatile service discipline dependencies, namely the three forms of N-policy.

4. PRELIMINARIES

In the following sections, we will be using some basic results from the first passage problem stated and developed by Abolnikov and Dshalalow [1].

As mentioned previously, we assume that inter-renewal times $t_n = \tau_n - \tau_{n-1}$, are characterized by their common Laplace-Stieltjes transform $e(\zeta) = E[e^{-\zeta t_n}] = \frac{\lambda}{\lambda + \zeta}$, $n = 1, 2, \ldots$, $Re(\zeta) \geq 0$. We also assume that the Poisson point process $\tau = \{\tau_n = t_0 + t_1 + \ldots + t_n; n \geq 0\}$ on $\mathbb{R}_+$ and the renewal process $S = \{S_n = X_0 + X_1 + \ldots + X_n; n \geq 0\}$ on $\{1, 2, \ldots\}$ are independent.

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

The following terminology was introduced in Abolnikov and Dshalalow [1] and we will use it throughout the remainder of this paper.

4.1 Definitions.

(i) For each n, the random variable $\nu_n = inf\{k \geq 0: S_k \geq N\}$ (defined in the previous section) is called the *index of the the first excess* (above level $N-1$).

(ii) The random variable S_{ν_n} is called the *level of the first excess* (above $N-1$).

(iii) The random variable τ_{ν_n} is the *first passage time* of S of level N.

Figure 1 is a graphic presentation of Model 1, where the levels are related as $r \leq N \leq R$. X_{i+1} is the batch arriving at instant τ_i. Let S_ν be the sum of X_i, $i = 0, 1, \ldots \nu$, where ν ($= \nu_n$, for brevity) is the smallest value at which S_ν is greater than level N. At instant τ_ν, the queue length S_ν exceeds the server capacity R; therefore, the server initiates the start-up process with start-up time $\xi_{\nu+1}$ followed by the service to be lasted $\sigma_{\nu+1}$. At the begin of that service time, $\tau_\nu + \xi_{\nu+1}$, the system takes a batch of $min\{S_\nu, R\}$, in our case, S_ν units for service. At the end of service, the system enters an idle state, since queue length Q_1, at instant T_1, becomes less than r.

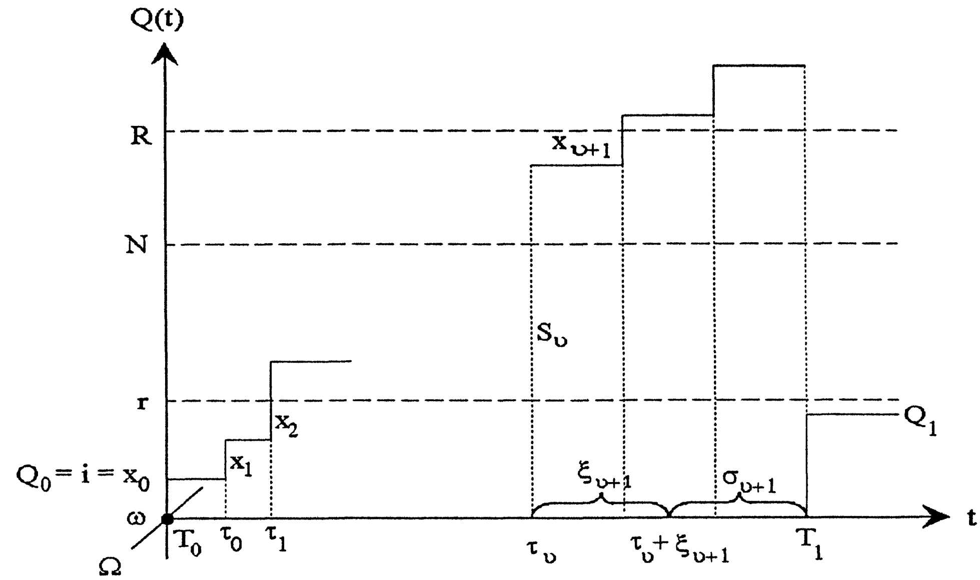

Figure 1 Model 1

Figure 2 is the graphic presentation of Model 2, where the levels are related as $r \leq R \leq N$. X_{i+1} is the batch size at instant τ_i. Let S_ν be the sum of X_i, $i = 0, 1, \ldots \nu$, where ν ($= \nu_n$, for brevity) is the smallest value at which S_ν is greater than N. At instant τ_ν, queue length S_ν exceeds the server capacity R; therefore, the server initiates the start-up process with start-up time $\xi_{\nu+1}$ followed by the service to be lasted $\sigma_{\nu+1}$. At the beginning of that service time $\tau_\nu + \xi_{\nu+1}$, the system takes a batch of $min\{S_\nu, R\}$, in our case, R units for service. At the end of service, the system keeps on being busy, since queue length Q_1, at instant T_1, is still greater than r.

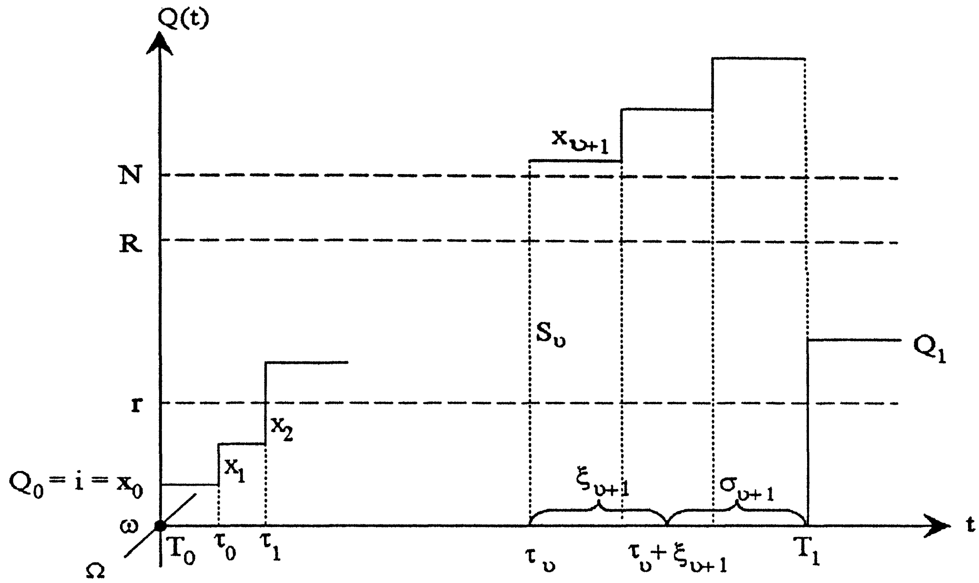

Figure 2 Model 2

Let the generating function of the first excess level be

$$G_{N-i}(z) = E^i[z^{S_{\nu_n}}] = E[z^{S_{\nu_n}} \mid Q_n = i] \ .$$

The following theorems were established in Abolnikov and Dshalalow [1] :

4.2 Theorem. *The generating function $\gamma_{N-i}(z)$ of the index of the first excess level satisfies the following formula:*

$$(4.2\text{a}) \quad \gamma_{N-i}(z) = \begin{cases} z\mathfrak{D}_x^{N-i-1}\left\{\dfrac{1-a(x)}{(1-x)(1-za(x))}\right\}, & i < N, \\ 1, & i \geq N, \end{cases}$$

where
$$\mathfrak{D}_x^k = \lim_{x \to 0} \frac{1}{k!}\frac{\partial^k}{\partial x^k}, \ k \geq 0,$$

with the mean value of the index of the first excess:

$$(4.2\text{b}) \quad \bar{\gamma}_{N-i} = \begin{cases} \mathfrak{D}_x^{N-i-1}\left\{\dfrac{1}{(1-x)[1-a(x)]}\right\}, & i < N, \\ 0, & i \geq N. \end{cases}$$

4.3 Theorem. *The generating function $G_{N-i}(z)$ of the first excess level can* *be determined from the following formula:*

$$\text{(4.3a)}\qquad G_{N-i}(z) = \begin{cases} z^N \mathfrak{D}_x^{N-i-1}\left\{\dfrac{a(z)-a(x)}{(z-x)(1-a(z))}\right\}, & i < N, \\ z^i, & i \geq N. \end{cases}$$

5. EMBEDDED PROCESS

The main objective of this section is the derivation of the stationary distribution of the discrete time parameter queueing process $\{Q_n\}$.

Model 1 ($r \leq N \leq R$) :

Let

$$A_i(z) = E^i[z^{Q_1}] = E[z^{Q_1} \mid Q_0 = i]\ , \qquad i \in \Psi = \{0,1,\ldots\}.$$

5.1 Proposition. *The generating function $A_i(z)$ of the ith row of transition probability matrix A can be determined from the following formulas:*

$$\text{(5.1a)}\qquad A_i(z) \;=\; K_i(z) z^{-R} H^R(G_{N-i})(z), \qquad i \in \Psi.$$

where the operator H^R is defined as

$$\text{(5.1b)}\qquad H^R(h)(z) = h(z) + \mathfrak{D}_y^{R-1}\left[\frac{z^{-R}h(y) - h(yz)}{1-y}\right], \quad i \in \Psi.$$

where h is a function analytic at zero, and

$$K_i(z) = \begin{cases} K(z)W(z), & i < r, \\ K(z), & i \geq r, \end{cases}$$

$$\text{(5.1c)}\qquad K(z) = \beta(\lambda - \lambda a(z)),$$

$$\text{(5.1d)}\qquad W(z) = \phi(\lambda - \lambda a(z)),$$

where

$$\mathfrak{D}_x^k = \begin{cases} \lim\limits_{x \to 0} \dfrac{1}{k!}\dfrac{\partial^k}{\partial x^k}, & k \geq 0, \\ 0\ , & k < 0. \end{cases}$$

and $\beta(\theta)$, $\phi(\theta)$, $(Re(\theta) \geq 0)$, are the Laplace-Stieltjes transforms of the corresponding general service time distribution function B and general start-up *time distribution function D.*

Proof: Let V_{n+1} denote the number of arrivals during the nth service. Since the input is an orderly stationary Poisson process, then:

$$E[z^{V_{n+1}}] = \beta(\lambda - \lambda a(z)) .$$

Similarly,

$$E[z^{W_{n+1}}] = \phi(\lambda - \lambda a(z)) .$$

Therefore, due to relation (3.3):

for $i \geq r$, $\quad A_i(z) = E[z^{(i-R)^+ + V_{n+1}} \mid Q_n = i] = z^{(i-R)^+} K(z)$,

and for $i < r$,
$$\begin{aligned} A_i(z) &= E[z^{(S_{\nu_n} - R)^+ + V_{n+1} + W_{n+1}} \mid Q_n = i] \\ &= E[z^{(S_{\nu_n} - R)^+} \mid Q_n = i] E[z^{V_{n+1}}] E[z^{W_{n+1}}] \\ &= E[z^{(S_{\nu_n} - R)^+} \mid Q_n = i] K(z) W(z) . \end{aligned}$$

From Abolnikov and Dshalalow [2], we have

$$E[z^{(S_{\nu_n} - R)^+} \mid Q_n = i] = z^{-R}\left\{G_{N-i}(z) - \mathfrak{D}_y^{R-1}\left[\frac{G_{N-i}(yz)}{1-y}\right]\right\} + \mathfrak{D}_y^{R-1}\left[\frac{G_{N-i}(y)}{1-y}\right] .$$

Therefore,

$$A_i(z) = K_i(z) z^{-R} H^R(G_{N-i})(z), \qquad i \in \Psi . \qquad \square$$

From relation (3.2) and our assumption about the input stream, we can conclude that $\{Q_n; n = 0, 1, \ldots\} \rightarrow \Psi = \{0, 1, \ldots\}.$, is a homogeneous, irreducible and aperiodic Markov chain.

We are interested in the transition probability matrix $A = (a_{ij}; i, j \in \Psi)$, where $a_{ij} = P^i\{Q_1 = j\} = P\{Q_1 = j \mid Q_0 = i\}$. This is of the form

$$A = \begin{array}{c} \\ 0 \\ 1 \\ \vdots \\ r-1 \\ r \\ r+1 \\ \vdots \\ R-1 \\ R \\ R+1 \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccccccc} 0 & 1 & 2 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{array} \\ \left(\begin{array}{ccccccccc} a_{00} & a_{01} & a_{02} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{10} & a_{11} & a_{12} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{r-1,0} & a_{r-1,1} & a_{r-1,2} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ f_0 & f_1 & f_2 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ f_0 & f_1 & f_2 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ f_0 & f_1 & f_2 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ f_0 & f_1 & f_2 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & f_0 & f_1 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{array} \right) \end{array}$$

where $A = (a_{i,j} : i, j \in \Psi : a_{i,j} = f_{j-i+R}, i \geq r, j \geq i-R;\ a_{i,j} = f_j, i \geq r, i \leq R;$ $a_{i,j} = 0, i \geq r, j < i-R)$,

$$a_{ij} = \begin{cases} P\{V_{n+1} = j-i+R\} = f_{j-i+R}\,, & \textit{if} \quad i \geq R, \\ P\{V_{n+1} = j\} = f_j\,, & \textit{if} \quad r \leq i < R. \end{cases}$$

The matrix A consists of two block matrices: the upper rectangular block, from row 0 to row $R-1$, with all positive elements, and the lower block, below row $R-1$, which is an upper triangular matrix. Matrix A is an $(R-1)$-homogeneous $\Delta_{R,R}$-matrix identified in Abolnikov/Dukhovny [3], where the general case of $\Delta_{m,n}$-matrices was introduced and studied.

According to Abolnikov and Dukhovny [3], the process $\{Q_n\}$ is recurrent-positive if and only if $\rho = \alpha\lambda b < R$. Therefore, for $\rho < R$, the Markov chain $\{Q_n\}$ is ergodic. Let $\boldsymbol{P} = (p_x\,; x \in \Psi)$ be the invariant probability measure of $\{Q_n\}$ and let $P(z)$ be the generating function of the components of vector $\boldsymbol{P}$.

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

(5.2a)
$$P(z) = \frac{K(z)[\sum_{i=0}^{r-1} p_i\{H^R(G_{N-i})(z)W(z) - z^i\} + \sum_{i=r}^{R-1} p_i\{z^R - z^i\}]}{z^R - K(z)},$$

where H^R *is given by* (5.1b), $K(z)$ *is given by* (5.1c), $W(z)$ *is given by* (5.1d). *Probabilities* $p_0, \ldots, p_{R-1}$ *form the unique solution of the following system of linear equations:*

$$(5.2b) \quad \frac{d^k}{dz^k}\{\textstyle\sum_{i=0}^{r-1} p_i\{H^R(G_{N-i})(z)W(z) - z^i\} + \sum_{i=r}^{R-1} p_i\{z^R - z^i\}]\}\Big|_{z=z_s} = 0,$$

$$k = 0, \ldots, k_s - 1, \; s = 1, \ldots, S,$$

$$(5.2c) \quad \textstyle\sum_{i=0}^{R-1}\left\{\alpha\lambda d + \alpha\overline{\gamma}_{N-i} + \mathfrak{D}_y^{R-1}\left[\frac{G_{N-i}(y)}{(1-y)^2}\right]\right\}p_i = R - \rho,$$

where z_s *are* $R-1$ *roots of* $z^R - K(z)$ *in* $\bar{B}(0,1)\backslash\{1\}$, *the closed unit disc (in the complex plane) centered at zero with deleted point* 1, *with their multiplicities* k_s *such that* $\sum_{s=1}^{S} k_s = R-1$.

Proof: According to Abolnikov and Dukhovny [3], an irreducible, aperiodic Markov chain Q_n with the transition probability matrix A (in the form of a $\Delta_{R,R}$-matrix), is recurrent-positive if and only if $\frac{d}{dz}A_i(z)\,|_{z=1} < \infty$, $i = 0, 1, \ldots, R$,

and $\quad \frac{d}{dz}K(z)\,|_{z=1} \leq R$.

Since $P(z) = \sum_{i=0}^{\infty} A_i(z)p_i$, and $A_i(z) = K_i(z)z^{-R}H^R(G_{N-i})(z)$, $i \in \Psi$, we have

$$\textstyle\sum_{i=0}^{R-1} z^i p_i + \sum_{i=R}^{\infty} z^i p_i = \sum_{i=0}^{R-1} K_i(z)z^{-R}H^R(G_{N-i})(z)p_i + \sum_{i=R}^{\infty} K(z)z^{i-R}p_i,$$

which yields

$$\textstyle\sum_{i=R}^{\infty} z^{i-R}p_i = \frac{\sum_{i=0}^{R-1}[K_i(z)z^{-R}H^R(G_{N-i})(z) - z^i]p_i}{z^R - K(z)}.$$

The left-hand side of this function is analytic in the open disc, $B(0,1)$, and continuous on its boundary, $\partial B(0,1)$.

By Abolnikov-Dukhovny's criterion, for $\rho < R$, $z^R - K(z)$ has exactly R zeros in $\bar{B}(0,1)$ (counting with their multiplicities); all zeros located on the boundary $\partial B(0,1)$ (including 1), are simple.

Now,

$$P(z) = z^{-R}\beta(\lambda - \lambda a(z))\left\{\textstyle\sum_{i=0}^{r-1} W(z)\left[z^N \mathfrak{D}_x^{N-i-1}\left[\frac{a(z) - a(x)}{(z-x)(1-a(x))}\right]\right.\right.$$

$$+\mathfrak{D}_{x,y}^{N-i-1,R-N-1}\left\{\frac{1}{(1-y)(1-a(x))}\left[z^R\left(\frac{a(z)-a(x)}{y-x}\right)-z^N\left(\frac{a(yz)-a(x)}{yz-x}\right)\right]\right\}\Bigg]p_i$$

$$+z^{-R}\beta(\lambda-\lambda a(z))\left\{z^R\sum_{i=r}^{R-1}p_i+\sum_{i=R}^{\infty}z^ip_i\right\}\Big\},$$

where $\quad \mathfrak{D}_{x,y}^{m,n}=\lim_{x\to 0,\,y\to 0}\frac{1}{m!n!}\frac{\partial^{m+n}}{\partial x^m\partial y^n},\qquad m,n\geq 0.$

By rearranging terms, we finally have

$$P(z)=\frac{K(z)[\sum_{i=0}^{r-1}[H^R(G_{N-i})(z)W(z)-z^i]p_i+\sum_{i=r}^{R-1}(z^R-z^i)p_i]}{z^R-K(z)},$$

where H^R is defined in (5.1b).

Since $z^R-K(z)$ has exactly R zeros in $\overline{B}(0,1)$, (5.2b) and (5.2c) represent a first order of simultaneous linear system of R equations and R unknowns which are the unknown probabilities $p_0, p_1,\ldots, p_{R-1}$. Therefore, we have a unique solution of the unknown probabilities $p_0, p_1,\ldots, p_{R-1}$.

By Proposition 5.1, we have $E^i[z^{Q_1}]=K_i(z)z^{-R}H^R(G_{N-i})(z)$, then formula
(5.2a) can be rewritten in the form

$$P(z)=\frac{\sum_{i=0}^{R-1}[z^RE^i[z^{Q_1}]-z^iK(z)]p_i}{z^R-K(z)}.$$

As in Abolnikov/Dshalalow [2], from $P(1)=1$, we have that

$$\sum_{i=0}^{R-1}\left\{\alpha\lambda d+\alpha\overline{\gamma}_{N-i}+\mathfrak{D}_y^{R-1}\left[\frac{G_{N-i}(y)}{(1-y)^2}\right]\right\}p_i=R-\rho.$$

This completes the proof of theorem 5.2. □

Model 2 ($r\leq R\leq N$) :

If $Q_n<r$, at the begin of $(n+1)$st service, the server capacity is R. Therefore, the system can now be described by (3.3).

With the same argument as that presented in Model 1, we have that

$$A_i(z)=\begin{cases}E^i[z^{S_{\nu_n}-R}]K(z)W(z), & i<N,\\ z^{i-R}K(z), & i\geq N.\end{cases}$$

For $i < r$, this yields

$$A_i(z) = z^{-R}\, E[z^{S_{\nu_n}}]K(z)W(z)$$
$$= z^{-R}\, G_{N-i}(z)K(z)W(z)\ ,$$

where

$$G_{N-i}(z) = \begin{cases} z^N \mathfrak{D}_x^{N-i-1}\left\{\dfrac{a(z)-a(x)}{(z-x)(1-a(z))}\right\}, & i < N\ , \\ z^i\ , & i \geq N. \end{cases}$$

For $i \geq r$, this yields

$$A_i(z) = z^{(i-R)^+} K(z)\ .$$

Now, we can restate the main theorem :

5.3 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:*

$$\text{(5.3a)}\qquad P(z) = \frac{K(z)[\sum_{i=0}^{r-1} p_i\{G_{N-i}(z)W(z) - z^i\} + \sum_{i=r}^{R-1} p_i\{z^R - z^i\}]}{z^R - K(z)}\ ,$$

where $K(z)$ satisfies (5.1c), $W(z)$ satisfies (5.1d).

Probabilities $p_0, \ldots, p_{R-1}$ form the unique solution of the following system of linear equations:

$$\frac{d^k}{dz^k}\left\{\sum_{i=0}^{r-1} p_i\{G_{N-i}(z)W(z) - z^i\} + \sum_{i=r}^{R-1} p_i\{z^R - z^i\}\right\}\Big|_{z=z_s} = 0\ ,$$
$$k = 0,\ldots,k_s - 1\ ,\ s = 1,\ldots,S,$$

$$\sum_{i=0}^{R-1}\{\alpha\lambda d + \alpha\overline{\gamma}_{N-i}\}p_i = R - \rho\ ,$$

where z_s are $R-1$ roots of $z^R - K(z)$ in $\bar{B}(0,1))\backslash\{1\}$ with their multiplicities k_s such that $\sum_{s=1}^{S} k_s = R - 1$.

Note that the case of $N = R$, the probability generating function $P(z)$ for Model 1 coincides with that of Model 2.

Model 3 ($r \leq N \leq R$) :

With the same argument as that presented in Model 1 and (3.4), we have that for $i \geq r$, $\quad A_i(z) = E[z^{(i-R)^+ + V_{n+1}} \mid Q_n = i] = z^{(i-R)^+} K(z)$,

and for $i < r$,
$$A_i(z) = E[z^{(S_{\nu_n} + W_{n+1} - R)^+ + V_{n+1}} \mid Q_n = i]$$
$$= E[z^{(S_{\nu_n} + W_{n+1} - R)^+} \mid Q_n = i] E[z^{V_{n+1}}]$$
$$= E[z^{(S_{\nu_n} + W_{n+1} - R)^+} \mid Q_n = i] K(z) .$$

From Abolnikov and Dshalalow [2], we have that

$$E[z^{(S_{\nu_n} + W_{n+1} - R)^+} \mid Q_n = i] = z^{-R} H^R(G_{N-i} W)(z) ,$$

where H^R was defined in (5.1b).

Therefore, we have

$$A_i(z) = \begin{cases} K(z) z^{-R} H^R(G_{N-i} W)(z), & i < r, \\ K(z) z^{(i-R)^+}, & i \geq r, \end{cases}$$

Now, we can restate the main theorem :

5.4 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:*

$$\text{(5.4a)} \quad P(z) = \frac{K(z)[\sum_{i=0}^{r-1} p_i \{H^R(G_{N-i} W)(z) - z^i\} + \sum_{i=r}^{R-1} p_i \{z^R - z^i\}]}{z^R - K(z)} ,$$

where $K(z)$ satisfies (5.1c), $W(z)$ satisfies (5.1d).

Probabilities $p_0, \ldots, p_{R-1}$ form the unique solution of the following system of linear equations:

$$\frac{d^k}{dz^k}\Big\{ \sum_{i=0}^{r-1} p_i \{H^R(G_{N-i} W)(z) - z^i\} + \sum_{i=r}^{R-1} p_i \{z^R - z^i\}\Big\}\Big|_{z = z_s} = 0 ,$$

$$k = 0, \ldots, k_s - 1 , \; s = 1, \ldots, S,$$

$$\sum_{i=0}^{R-1} \left\{ \alpha\lambda d + \alpha \overline{\gamma}_{N-i} + \mathfrak{D}_y^{R-1}\left[\frac{G_{N-i}(y) W(y)}{(1-y)^2}\right]\right\} p_i = R - \rho ,$$

where z_s are $R-1$ roots of $z^R - K(z)$ in $\bar{B}(0,1))\backslash\{1\}$ with their multiplicities k_s such that $\sum_{s=1}^{S} k_s = R - 1$.

Note that in the case of where the start-up time requirement is dropped, and $r = N$, the probability generating function $P(z)$ for Model 3 coincides with that of the model developed by Abolnikov and Dshalalow [2].

6. SPECIAL CASES AND NUMERICAL EXAMPLES

In the following, we discuss some special cases of Model 1.

6.1 Special Case : Let us drop the N-policy; i.e., $N = r$. The probability generating function can be written as :

$$P(z) = \frac{\beta(\lambda - \lambda a(z))}{z^R - \beta(\lambda - \lambda a(z))}$$

$$\cdot \Big[\sum_{i=0}^{r-1} \left\{ \left[G_{r-i}(z) - z^R \mathfrak{D}_y^{R-1} \left\{ \frac{G_{r-i}(y) - z^{-R} G_{r-i}(yz)}{1-y} \right\} \right] \phi(\lambda - \lambda a(z)) - z^i \right\} p_i$$

$$+ \sum_{i=r}^{R-1} (z^R - z^i) p_i \Big] .$$

Furthermore, if we drop the start-up time condition ($\phi(\lambda - \lambda a(z)) = 1$), we get

$$P(z) = \frac{\beta(\lambda - \lambda a(z))}{z^R - \beta(\lambda - \lambda a(z))}$$

$$\cdot \Big[\sum_{i=0}^{r-1} \left\{ \left[G_{r-i}(z) - z^R \mathfrak{D}_y^{R-1} \left\{ \frac{G_{r-i}(y) - z^{-R} G_{r-i}(yz)}{1-y} \right\} \right] - z^i \right\} p_i$$

$$+ \sum_{i=r}^{R-1} (z^R - z^i) p_i \Big] .$$

This result represents a special case of $M^x/G^{r,R}/1$ studied in Abolnikov/Dshalalow [2], where in our case the modulation and state dependency is dropped. (For details see the notion of modulation in [2] or our notice on modulation briefly mentioned at the end of section 3).

6.2 Special Case : In Model 1, we drop the bilevel service discipline by letting $r = R = 1$, but retain the start-up time parameter. The probability generating function then can be reduced to :

$$P(z) = \frac{\beta(\lambda - \lambda a(z))}{z - \beta(\lambda - \lambda a(z))} \left\{ z^N \mathfrak{D}_x^{N-1} \left[\frac{a(z) - a(x)}{(z-x)(1-a(x))} \right] \phi(\lambda - \lambda a(z)) - 1 \right\} p_0 ,$$

where $p_0 = \dfrac{1 - \alpha\lambda b}{\alpha(\overline{\gamma}_N + \lambda d)}$.

Furthermore, if the bulk arrival condition is dropped, i.e., $a(z) = z$, the probability generating function can be simplified to :

$$P(z) = \frac{\beta(\lambda - \lambda z)}{z - \beta(\lambda - \lambda z)} \left\{ z^N \phi(\lambda - \lambda z) - 1 \right\} p_0 ,$$

with $p_0 = \dfrac{1 - \lambda b}{N + \lambda d}$.

This result agrees with the M/G/1 queueing system studied by Medhi and Templeton [16].

In the following numerical examples, for simplicity of calculations the start-up time parameter is dropped in Model 1. (Note that the start-up time parameter affects only the multiplier, $W(z)$, in the first term of the formula shown in (5.2a), (5.3a), and (5.4a).)

6.3 Example:

In Model 1 ($r \le N \le R$), it is understood that as long as queue length is less than r, the system under study will become idle. In order to keep system idle time minimal, the "system turned-off probability", which is sum of the probability of queue length from 0 through $r-1$, needs to be calculated and compared under various values of r and N. Select parameters r and N in such a combination that the smallest possible "system turned-off probability" under steady state can be achieved. Assume that the bulk arrival groups have a geometric distribution with parameter $p = 0.3$ and assume that service time is exponentially distributed with rate $b = 0.2$. For a numerical demonstration, we take $R = 6$, and have r and N run from 0 to 5 individually.

Thus,

$$a(z) = \frac{pz}{1 - qz} .$$

The Laplace-Stieltjes transform of service time distribution function with rate $\mu = \frac{1}{b}$ is

$$\beta(s) = \frac{\mu}{\mu + s} .$$

Then, $\beta(\lambda - \lambda a(z)) = \dfrac{1}{1 + \rho - \rho a(z)}$, where $\rho = \frac{\lambda}{\mu}$, is the system intensity.

Thus, after some algebra, we have

$$P(z) = \frac{1}{-(\rho+q)z^6 + pz^5 + \ldots + pz + 1}$$

$$\cdot\left\{\sum_{i=0}^{r-1} z^i[(q^{6-N-1} - q)z^{6-i} + p\sum_{k=1}^{6-i-1} z^k + 1]p_i\right.$$

$$\left. + (1-qz)[\sum_{i=r}^{6-1} z^i(\sum_{k=0}^{6-i-1} z^k)p_i]\right\}.$$

With this formula, we can start the numerical evaluation. First, we set up the linear system:
(5.2c) and (5.2b) yield

$$\text{(6.3.1a)} \qquad \sum_{i=0}^{r-1}[q^{6-N+1} + (6-i)p]p_i + p\sum_{i=r}^{6-1}(6-i)p_i = 6p - \rho\ ,$$

and

$$\frac{d^k}{dz^k}[\sum_{i=0}^{r-1} z^i[(q^{6-N-1} - q)z^{6-i} + p\sum_{k=1}^{6-i-1} z^k + 1]p_i$$

$$+ (1-qz)[\sum_{i=r}^{6-1} z^i[\sum_{k=0}^{6-i-1} z^k]p_i]\Big|_{z=z_s} = 0\ ,$$

for $k = 0, 1, 2, \ldots, k_s - 1$, and $s = 1, \ldots, S$,

where z_s are the 5 roots of $-(\rho+q)z^6 + pz^5 + \ldots + pz + 1$ in the region $\bar{B}(0,1)\backslash\{1\}$, and with their multiplicities k_s such that $\sum_{s=0}^{S-1} k_s = 5$.

With the above assumptions, according Dukhovny [13], the equation

$$-(\rho+q)z^6 + pz^5 + \ldots + pz + 1 = 0$$

has no multiple roots in the region of $\bar{B}(0,1)\backslash\{1\}$. Therefore, (5.2b) can further be reduced to

$$\text{(6.3.1b)} \quad \sum_{i=0}^{r-1} z_j^i[(q^{6-N-1} - q)z^{6-i} + p\sum_{k=1}^{6-i-1} z_j^k + 1]p_i$$

$$+ (1-qz_j)[\sum_{i=r}^{6-1} z_j^i[\sum_{k=0}^{6-i-1} z_j^k]p_i = 0\ ,$$

where z_j, $j = 1, 2, \ldots, 5$ are the 5 roots of the equation

$$-(\rho+q)z^6 + pz^5 + \ldots + pz + 1 = 0\ .$$

Note here, that by eliminating the common factor from both the numerator and denominator of the probability generating function $P(z)$, a single root $z_0 \notin \bar{B}(1,0)\backslash\{1\}$ as a byproduct of multiplication was added to the equation. In most cases, $|z_0| > 1$.

Combining (6.3.1a) and (6.3.1b), we can find p_i, $i = 0, 1, 2, \ldots, 5$. For example, let $R = 6$, $N = 4$, $r = 2$, and z_j, $j = 1, 2, 3, 4, 5$, be the single roots of

$$-(p+q)z^6 + pz^5 + pz^4 + pz^3 + pz^2 + pz + 1 = 0 .$$

The equation (6.3.1a) can be written explicitly as follows

(6.3.2a) $(q^3 + 6p)p_0 + (q^3 + 5p)p_1 + 4pp_2 + 3pp_3 + 2pp_4 + pp_5 = 6p - \rho$,

and the equation (6.3.1b) can be written explicitly as follows

(6.3.2b)
$$\begin{aligned}
&[(q^3 - q)z_j^6 + pz_j^5 + pz_j^4 + pz_j^3 + pz_j^2 + pz_j + 1]p_0 \\
&+ [(q^3 - q)z_j^5 + pz_j^4 + pz_j^3 + pz_j^2 + pz_j + 1]z_j p_1 \\
&+ [(1 - qz_j)(z_j^3 + z_j^2 + z_j + 1)z_j^2]p_2 \\
&+ [(1 - qz_j)(z_j^2 + z_j + 1)z_j^3]p_3 \\
&+ [(1 - qz_j)(z_j + 1)z_j^4]p_4 \\
&+ [(1 - qz_j)z_j^5]p_5 = 0 , \qquad j = 1, 2, 3, 4, 5 .
\end{aligned}$$

Thus, the probabilities p_0, p_1, p_2, p_3, p_4, p_5, can be found by solving linear systems (6.3.2.a) and (6.3.2.b).

Table 1 summarizes the numerical evaluation of probabilities for various control parameters. These parameters are:

Input arrival Poisson distribution parameter $\lambda = 5.0$,

Service time exponential distribution parameter $\mu = 5.0$,

System intensity $\rho = 1.0$,

Bulk arrival (geometric distribution) parameter $p = 0.3$,

Server capacity $R = 6$.

The polynomial in the denominator is :

$$z^R - K(z) = -1.70\ z^6 + 0.30\ z^5 + 0.30\ z^4 + 0.30\ z^3 + 0.30\ z^2 + 0.30\ z + 1 .$$

Roots of the above polynomial are:

$z_1 = 0.4245 - 0.7776i$,

$z_2 = 0.4245 + 0.7776i$,

$z_3 = -0.4468 + 0.7583i$,

$z_4 = -0.4468 - 0.7583i$,

$z_5 = -0.8793$.

$z_6 = 1.1003$, an artificially acquired root due to multiplication. (as previously noted.)

Table 1 provides information to show that when server capacity stays the same, say $R = 6$, if level r is fixed and level N increases, the total system idle probability becomes smaller. But when level N fixed and level r decreases, the total system idle probability also decreases. Thus, in order to achieve the optimal solution of the smallest possible system turned-off probability, level N should be set as high as possible and level r should be set as low as possible.

6.4 Example:

The following example preserves the same conditions of example 6.3, except for the service time distribution, which now is 2-Erlang with parameter μ.

$$B_1(x) = \begin{cases} \frac{2\mu(2\mu x)}{1!}e^{-2\mu x}, & x > 0, \\ 0\ , & x \leq 0, \end{cases}$$

where $\mu = \frac{2}{b}$,
with its Laplace-Stieltjes transform $\beta_1(s) = \left(\frac{\mu}{s+\mu}\right)^2$.

Now, $\beta_1(\lambda - \lambda a(z)) = \left(\frac{1}{1+\rho-\rho a(z)}\right)^2$, where $\rho = \frac{\lambda}{\mu}$.

With similar calculation, we obtain the following formula:

(5.2c) and (5.2.b) yield

(6.4.1a)

$$(1-q)\sum_{i=0}^{r-1}[q^{R-N+1} + (R-i)p]p_i + (1-q)^2\sum_{i=r}^{R-1}(R-i)p_i = p(pR-2\rho),$$

(6.4.1b)

$$P(z) = \frac{1}{-(\rho+q)^2z^{R+1} - (q+\rho)(1+p+\rho)z^R + p^2z^{R-1} + \ldots + p^2z^2 + (p-q)z + 1} \cdot \Big\{(1-qz)\sum_{i=0}^{r-1}z^i[(q^{R-N-1}-q)z^{R-i} + p\sum_{k=1}^{R-i-1}z^k + 1]p_i + (1-qz)^2\sum_{i=r}^{R-1}z^ip_i\sum_{k=0}^{R-i-1}z^k\Big\}$$

Combining (6.4.1a) and (6.4.1b), we can find the probabilities of queue length equals i, p_i, where $i = 0, 1, 2, \ldots, R-1$.

Let $R = 6$, $N = 4$, $r = 2$, and according Dukhovny [13], the equation

$$-(\rho+q)^2z^7 - (q+\rho)(1+p+\rho)z^6 + p^2z^5 + \ldots + p^2z^2 + (p-q)z + 1 = 0\ .$$

has no multiple roots in the region of $\overline{B}(0,1)\backslash\{1\}$.

(6.4.1a) can be written explicitly as follows

(6.4.2a)

$$(1-q)(q^3+6p)p_0 + (q^3+5p)p_1 + (1-q)^2[4p_2+3p_3+2p_4+p_5] = p[6p-2\rho] .$$

And let z_j, $j=1, 2, 3, 4, 5$ be the single roots of

$$-(\rho+q)^2 z^7 - (q+\rho)(1+p+\rho)z^6 + p^2 z^5 + \ldots + p^2 z^2 + (p-q)z + 1 = 0 ,$$

thus, (6.4.1b) can be written explicitly as follows

(6.4.2b)
$$\begin{aligned}
&(1-qz_j)[(q^3-q)z_j^6 + pz_j^5 + pz_j^4 + pz_j^3 + pz_j^2 + pz_j + 1]p_0 \\
&+ (1-qz_j)\,[(q^3-q)z_j^5 + pz_j^4 + pz_j^3 + pz_j^2 + pz_j + 1]z_j p_1 \\
&+ (1-qz_j)^2[(z_j^3 + z_j^2 + z_j + 1)z_j^2]p_2 \\
&+ (1-qz_j)^2[(z_j^2 + z_j + 1)z_j^3]p_3 \\
&+ (1-qz_j)^2[(z_j + 1)z_j^4]p_4 \\
&+ (1-qz_j)^2 z_j^5 p_5 = 0 , \qquad j = 1, 2, 3, 4, 5 .
\end{aligned}$$

Thus, the probabilities p_0, p_1, p_2, p_3, p_4, p_5, can be found by solving linear systems (6.4.2.a) and (6.4.2.b).

Table 2 summarizes the numerical evaluation of probabilities from various control policies. The parameters are :

Input arrival Poisson distribution parameter $\lambda = 5.0$
Service time 2-Erlang distribution parameter $\mu = 10.0$
System intensity $\rho = 0.5$
Bulk arrival (geometric distribution) parameter $p = 0.3$,
Server capacity $R = 6$.

The polynomial in the denominator is :

$$z^6 - K(z) = 1.44\, z^7 - 2.16\, z^6 + 0.09\, z^5 + 0.090\, z^4 + 0.09\, z^3 + 0.090\, z^2 - 0.040\, z + 1 .$$

Roots of the above polynomial are :

$z_1 = 0.4101 + 0.7650i$,
$z_2 = 0.4101 - 0.7650i$,
$z_3 = -0.4387 + 0.7394i$,
$z_4 = -0.4387 - 0.7394i$,
$z_5 = -0.8585$.
$z_6 = 1.2883$, an artificially acquired root due to multiplication (as previously noted.)
$z_7 = 1.1274$, an artificially acquired root due to multiplication (as previously noted.)

Table 2 also indicates that when server capacity stays the same, $R = 6$, if level N increases, the total system idle probability becomes smaller. However, under the same assumption, when level r decreases, so does the total system idle probability. Therefore, similar to Example 6.3, in order to achieve the smallest possible system turned-off probability under steady state, one should set level r as low and level N as high as conditions permit.

ACKNOWLEDGEMENTS

The author wishes to express his appreciation to Dr. Heinz H. Schreiber, Professor of Electrical Engineering, Florida Institute of Technology, and Deputy Director of Systems Engineering, Grumman Melbourne Systems, for his encouragement and technical support, and to Dr. Jan Herndon of Grumman Melbourne Systems who provided a very careful proofreading of the paper and gave a number of useful suggestions.

The author also wishes to thank the anonymous referee whose many insightful suggestions and constructive comments have considerably enhanced the presentation of this paper.

REFERENCES

[1] Abolnikov, L.M. and Dshalalow, J.H., A first passage problem and its application to the analysis of a class of stochastic models, *JAMSA*, 5, No. 1, 83-98, 1992.

[2] Abolnikov, L.M. and Dshalalow, J.H., On a multilevel controlled bulk queueing system $M^x/G^{r,R}/1$, *JAMSA*, 5, No. 3, 237-260, 1992.

[3] Abolnikov L.M. and Dukhovny A.M., Markov chains with transition delta-matrix: ergodicity conditions, invariant probability measures and applications, *JAMSA*, 4, No. 4, 335-355, 1991.

[4] Arora, R.T., and Tuteja, R.K., Random N-policy for a single server queue with linear cost structure, *Soochow J. Math.*, 17, No. 1, 33-53, 1991.

[5] Baker, K.R., A note on operating policies for the queue M/M/1 with exponential startups, *INFOR* 11, 71-72, 1973.

[6] Bohm, W., A transient analysis of M/G/1 queues with N-policy, *Statist. Hefte*, 33, No. 2, 151-157, 1992.

[7] Bohm, W., The transient solution of M/M/1 queues under (M,N)-policy. A combinatorial approach, *J. Statist. Plann. Inference*, 34, No.1, 23-33, 1993.

[8] Borthakur, A., Medhi, J., and Gohain, R., Poisson input queueing with startup time

and under control operating policy, *Computers Ops Res.* 14, 33-40, 1987.

[9] Chaudhry, M.L. and Templeton, J.G.C., *A first course in Bulk Queues,* John Wiley and Sons, New York, 1983.

[10] Chitkara, A.K., and Kumar, A., N-policy for an Erlangian input system with reorientation period, *Investigacion Oper. 7,* No. 1, 57-66, 1986.

[11] Cooper, R.B., *Introduction to Queueing Theory,* North Holland, New York, 1981.

[12] Daigle, J.N., *Queueing Theory for Telecommunications,* Addison Wesley, 1992.

[13] Dukhovny, A.M., Some remarks on multiple roots of characteristic equations (to appear in Stoch. Models).

[14] Heyman, D.P., Optimum operating policies for M/G/1 queueing systems, *Ops Res.* 16, 362-382, 1968.

[15] Igaki, N., Exponential two server queue with N-policy and general vacations, *Queueing Systems, Theory and Appl.* 10, No. 4, 1992.

[16] Medhi, J., and Templeton, J.G.C., A Poisson input queue under N-policy and with a general start-up time, *Computers Ops Res.* Vol. 19, No. 1, 35-41, 1992.

[17] Rosner, R.D., *Packet Switching,* Lifetime Learning Publications, Belmont, CA. 1982.

[18] Talman, A.J.J., A simple proof of the optimality of the best N-policy in the M/G/1 queueing control problem with removable server, *Statist. Neerlandica,* 33, No. 3, 143-150, 1979.

[19] Tian, N.S., The N-policy stochastic service system GI/M/1, *Math. Practice Theory,* No. 1, 27-33, 1992.

Control Policy		Probabilities						
r	N	p_0	p_1	p_2	p_3	p_4	p_5	Turned off*
5	5	0.1585	0.0767	0.0697	0.0634	0.0576	0.0523	0.4259
4	5	0.1742	0.0753	0.0484	0.0622	0.0565	0.0514	0.3800
4	4	0.2072	0.0723	0.0657	0.0597	0.0543	0.0493	0.4048
3	5	0.1907	0.0738	0.0670	0.0609	0.0554	0.0503	0.3315
3	4	0.2191	0.0712	0.0647	0.0588	0.0534	0.0486	0.3549
3	3	0.2414	0.0691	0.0628	0.0571	0.0519	0.0472	0.3734
2	5	0.2082	0.0722	0.0656	0.0596	0.0542	0.0492	0.2804
2	4	0.2318	0.0700	0.0636	0.0578	0.0526	0.0478	0.3018
2	3	0.2505	0.0683	0.0621	0.0564	0.0513	0.0466	0.3188
2	2	0.2650	0.0670	0.0609	0.0553	0.0503	0.0457	0.3320
1	5	0.2265	0.0705	0.0641	0.0582	0.0529	0.0481	0.2265
1	4	0.2453	0.0688	0.0625	0.0568	0.0516	0.0469	0.2453
1	3	0.2603	0.0674	0.0613	0.0557	0.0506	0.0460	0.2603
1	2	0.2720	0.0664	0.0603	0.0548	0.0498	0.0453	0.2720
1	1	0.2808	0.0656	0.0596	0.0541	0.0492	0.0447	0.2808

Turned-off* represents the sum of the probability of queue length from 0 to r-1

Table 1.

Control Policy		Probability						
r	N	p_0	p_1	p_2	p_3	p_4	p_5	Turned off*
5	5	0.0551	0.0296	0.0277	0.0257	0.0237	0.0217	0.1619
4	5	0.0618	0.0298	0.0278	0.0257	0.0237	0.0216	0.1451
4	4	0.0694	0.0279	0.0258	0.0238	0.0218	0.0199	0.1469
3	5	0.0706	0.0307	0.0286	0.0264	0.0242	0.0221	0.1299
3	4	0.0769	0.0289	0.0267	0.0246	0.0225	0.0204	0.1325
3	3	0.0815	0.0276	0.0254	0.0233	0.0212	0.0192	0.1344
2	5	0.0829	0.0326	0.0302	0.0278	0.0254	0.0232	0.1154
2	4	0.0878	0.0308	0.0285	0.0261	0.0238	0.0216	0.1186
2	3	0.0914	0.0296	0.0272	0.0249	0.0226	0.0205	0.1210
2	2	0.0940	0.0287	0.0263	0.0240	0.0218	0.0197	0.1227
1	5	0.1009	0.0359	0.0332	0.0304	0.0278	0.0253	0.1009
1	4	0.1044	0.0343	0.0315	0.0288	0.0263	0.0238	0.1044
1	3	0.1069	0.0330	0.0303	0.0276	0.0251	0.0227	0.1069
1	2	0.1088	0.0321	0.0294	0.0268	0.0243	0.0220	0.1088
1	1	0.1102	0.0315	0.0287	0.0261	0.0237	0.0214	0.1101

Turned-off* represents the sum of the probablity of queue length from 0 to r-1

Table 2.