

## A DISCRETE SINGLE SERVER QUEUE WITH MARKOVIAN ARRIVALS AND PHASE TYPE GROUP SERVICES<sup>1</sup>

ATTAHIRU SULE ALFA and K. LAURIE DOLHUN

*University of Manitoba  
Department of Mechanical and Industrial Engineering  
Winnipeg, Manitoba, Canada R3T 5V6*

S. CHAKRAVARTHY

*GMI Engineering and Management Institute  
Department of Science and Mathematics  
Flint, MI 48504-4898 USA*

(Received October, 1994; Revised March, 1995)

### ABSTRACT

We consider a single-server discrete queueing system in which arrivals occur according to a Markovian arrival process. Service is provided in groups of size no more than  $M$  customers. The service times are assumed to follow a discrete phase type distribution, whose representation may depend on the group size. Under a probabilistic service rule, which depends on the number of customers waiting in the queue, this system is studied as a Markov process. This type of queueing system is encountered in the operations of an automatic storage retrieval system. The steady-state probability vector is shown to be of (modified) matrix-geometric type. Efficient algorithmic procedures for the computation of the rate matrix, steady-state probability vector, and some important system performance measures are developed. The steady-state waiting time distribution is derived explicitly. Some numerical examples are presented.

**Key words:** Markovian Arrival Process, Phase Type Distributions, Matrix Geometric Method, Group Service.

**AMS (MOS) subject classifications:** Primary: 60K20, 60K25, 90B22; Secondary: 60J27, 60K05 60K15.

### 1. Introduction

We consider a discrete time single-server queue in which customers arrive according to a Markovian Arrival Process (MAP). Service is provided in groups of at least one but not more than  $M$  customers at a time, where  $M$ , a fixed constant, gives the upper limit for the server capacity. The service times of each group are assumed to follow a discrete phase type distribution whose parameters may depend on the group size. When the server becomes free, that person starts to serve the first  $M$  of the waiting customers if the number waiting is at least  $M$ ; otherwise, he or she initiates a service at the next arrival for group of service size  $i$ , with probability  $\theta_i$ ,  $1 \leq i \leq M$ . With probability  $1 - \theta_i$ , the server remains idle. Note that  $\theta_M = 1$ .

---

<sup>1</sup>This research was supported in part by Grant No. OGP0006584 from the Natural Sciences and Engineering Research Council of Canada to A.S. Alfa and National Science Foundation Grant No. DDM-9313283 to S. Chakravarty.

The set of probabilities  $\theta_i$  plays an interesting role in this type of system. The functional form of  $\theta_i$  is usually non-decreasing in most real life situations and the important aspect that affects the system performance is whether it is a non-decreasing convex, concave or step function. For example, a convex function reflects the behavior of a very conservative server who is not overly anxious to start service until the number of the customers waiting is very close to  $M$ . A concave function reflects the behavior of an aggressive server who is anxious to start service without keeping customers waiting for too long. In both cases, the server uses judgment to decide when to start serving groups less than  $M$ . On the other hand, a step function refers to a deterministic case where the server knows when and when not to start service. Later in the paper we investigate the impacts of these functional forms of  $\theta_i$  on the performance measures of the system.

This type of queuing system is encountered in an automated storage and retrieval system where the material handling system has a capacity of carrying a maximum of  $M$  items. A decision has to be made whether to proceed and attend to the waiting items when there is less than  $M$  waiting. Usually the service time involves the long haul to the appropriate region in the racks plus the time per item to be placed in rack positions. Hence, the service times may be dependent on the group size.

Another application of this model is in the package delivery service by a courier company. Packages to be shipped by such companies are often time sensitive. For packages with common destination, the company would like these in batches of reasonable size. The company may thus decide, for economic reasons, to have a minimum number of packages to form a batch before they are shipped, and a maximum number at which the packages must be shipped. In between those two numbers the dispatcher has to decide when to ship the packages. In this example the vehicles are the servers and the packages are the customers.

Neuts and Nadarajan [10] considered problems of this class with many servers. They assume Poisson arrivals and exponential services with parameters independent of the group size. In their model a server is not allowed to serve less than  $a$ , and not more than  $b$  customers. Sim and Templeton [11] considered the same type of model but allow a probability distribution for the batch size (between  $a$  and  $b$ ). Later Chakravarty and Alfa [1] extended that model to include a Markovian arrival process and also introduced a probability  $\theta_i$  of serving the  $i$  waiting customers ( $a \leq i \leq b$ ).

Earlier Medhi and Borthakur [6] considered the case involving two servers and general bulk service. Medhi [5] later extended this model to the case involving  $c$ -servers.

## 2. The Model

The arrival process is assumed to be a MAP described by two substochastic matrices  $D_0$  and  $D_1$  of dimension  $n$ . The MAP was introduced in Neuts [9] as a generalization of the Poisson process which is well suited for matrix analytic and numerical investigations. In conjunction with the research reported by Lucantoni, Meier-Hellstern and Neuts [3], Lucantoni [2] suggested a convenient notation which is better suited for a general discussion than the one which was originally used.

A highly accessible discussion of the MAP, with many examples, may be found in Lucantoni [2]. A partly expository paper discussing how the MAP can be used qualitatively to model point processes with certain "bursty" features is given by Neuts [7]. The following is a brief informal description of the discrete MAP, (DMAP), which should be adequate for the purposes of this paper. Let  $D$ , ( $= D_0 + D_1$ ) be an irreducible stochastic matrix of order  $m$  and let  $D_0$  and  $D_1$  be two substochastic matrices whose sum is  $D$  and such that the matrix  $I - D_0$  is nonsingular.

At any transition from state  $i$  to state  $j$  of the Markov chain with transition probability matrix  $D$ , that transition is an arrival epoch with conditional probability  $(D_1)_{ij}[D_{ij}]^{-1}$  and is not an arrival epoch with the complementary conditional probability  $(D_0)_{ij}[D_{ij}]^{-1}$ . Equivalently, we may consider the discrete-time Markov renewal process embedded at arrival epochs and with transition probabilities defined by the sequence of matrices  $Q(k) = [D_0]^{k-1}D_1$ , for  $k \geq 1$ .

The DMAP is a discrete-time point process generated by the transition epochs of the above Markov renewal process. The simplest DMAP is the Bernoulli process for which the matrices  $D_0$  and  $D_1$  are scalars, respectively given by,  $1 - p$  and  $p$ . The point process is then the sequence of epochs of successes in independent Bernoulli trials.

The stationary vector  $\pi$  of the Markov chain described by  $D_0 + D_1$  satisfies the equation:

$$\pi(D_0 + D_1) = \pi, \quad \pi e = 1 \tag{1}$$

where  $e$  is a column vector given  $e = [1, 1, \dots, 1]'$ , where  $'$  denotes the transpose operation and  $\pi D_1 e$  is the probability that, in the stationary version of the arrival process, there is an arrival at an arbitrary time point. Correspondingly,  $\lambda^* = \pi D_1 e$  is the expected number of arrivals per unit of time and is also referred to as the fundamental rate of the MAP.

Service is provided in groups of sizes  $i$ ,  $1 \leq i \leq M$ , with probability  $\theta_i$ . The service for a group of size  $i$  is of phase type with representation  $(\beta_i, S_i)$  of order  $n_i$ . For later use, let  $S_i^0$  be such that  $S_i e + S_i^0 = e$ .

Note that  $(\mu_i^*)^{-1} = \beta_i(I - S_i)^{-1}e$ , where  $\mu_i^*$  is the service rate of the system when a group of  $i$  customers is served.

Let us define some notations that will be used in the sequel. The symbol  $\otimes$  stands for the Kronecker product. That is,  $X \otimes Y$  stands for the matrix made up of blocks  $X_{ij}Y$ . For more details about Kronecker products see Marcus and Minc [4].

Suppose  $U$  and  $V$  are square matrices of order  $m$  and  $n$  respectively. Then, the direct sum of  $U$  and  $V$ , denoted by  $U \dot{+} V = W$ , is an  $(n + m) \times (n + m)$  matrix with

$$W_{ij} = U_{ij}(i, j = 1, \dots, m) \tag{2}$$

$$W_{m+i, m+j} = V_{ij}(i, j = 1, \dots, n) \tag{3}$$

$$W_{ij} = 0, \text{ otherwise.} \tag{4}$$

In a partitioned form we write

$$W = \begin{bmatrix} U & 0 \\ 0 & V \end{bmatrix}.$$

Hence

$$U_1 \dot{+} U_2 \dot{+} \dots \dot{+} U_n = \begin{bmatrix} U_1 & 0 & \dots & \dots & 0 \\ 0 & U_2 & 0 & \dots & 0 \\ 0 & 0 & U_3 & \vdots & \vdots \\ \vdots & \vdots & \dots & \ddots & 0 \\ 0 & 0 & \dots & \dots & U_n \end{bmatrix}.$$

The variable  $\nu$  is defined as  $\nu = n_1 + n_2 + n_3 + \dots + n_M$  and  $e_j$ , is a column vector of appropriate dimension with 1 in the  $j^{th}$  position and 0 elsewhere.

### 3. The Markov Chain

The system described in Section 2 can be studied as a Markov chain. The state space for this system can be described by  $\Delta = \{(j, k), 0 \leq j \leq M - 1, 1 \leq k \leq n\} \cup \{(i, j_1, j_2, k, j_3), i \geq 0, 0 \leq j_1 \leq M - 1, 1 \leq j_2 \leq M, 1 \leq k \leq n, 1 \leq j_3 \leq n_{j_2}\}$ .

The states and their description are given in Table 1 below.

Table 1 - States and their description

State	Description
$(j, k)$ $0 \leq j \leq M - 1, 1 \leq k \leq n$	The server is idle with $j$ customers waiting and the arrival process is in phase $k$ .
$(i, j_1, j_2, k, j_3)$ $i \geq 0, 0 \leq j_1 \leq M - 1, 1 \leq j_2 \leq M$ $1 \leq k \leq n, 1 \leq j_3 \leq n_{j_2}$	The server is busy serving a group of $j_2$ customers with their service in phase $j_3$ , $(i \times M) + j_1$ , customers are waiting and the arrival process is in phase $k$ .  Note that the number of customers waiting is organized in $i$ groups of equal sizes $M$ and the remaining $j_1$ less than $M$ . This is because the server will serve in groups of size $M$ whenever there is at least that number waiting.

The transition matrix of this Markov chain is given by:

$$P = \begin{bmatrix} B_{00} & B_{01} & 0 & 0 & 0 & 0 & \dots & \dots & \dots \\ B_{10} & B_{11} & A_0 & 0 & 0 & 0 & \dots & \dots & \dots \\ 0 & A_2 & A_1 & A_0 & 0 & 0 & \dots & \dots & \dots \\ 0 & 0 & A_2 & A_1 & A_0 & 0 & \dots & \dots & \dots \\ 0 & 0 & 0 & A_2 & A_1 & A_0 & \dots & \dots & \dots \\ \vdots & \vdots \end{bmatrix}$$

where

$$B_{00} = \Gamma_0 \quad \text{is a square matrix of dimension } nM \tag{5}$$

$$B_{01} = e'_1 \otimes \Psi_0 \quad \text{is of dimension } (nM \times nM \nu) \tag{6}$$

$$B_{10} = \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \\ \vdots \\ \Gamma_M \end{bmatrix} \quad \text{is of dimension } (nM \nu \times nM) \tag{7}$$

$B_{11}, A_0, A_1$ , and  $A_2$  are square matrices of dimension  $nM \nu$  given by

$$B_{11} = \begin{bmatrix} \Psi_1 & G_0 & 0 & 0 & 0 & \dots & \dots & 0 & 0 \\ \Psi_2 & G_1 & G_0 & 0 & 0 & \dots & \dots & 0 & 0 \\ \Psi_3 & 0 & G_1 & G_0 & 0 & \dots & \dots & 0 & 0 \\ \Psi_4 & 0 & 0 & G_1 & G_0 & \dots & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \dots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \dots & \vdots & \vdots \\ \Psi_{M-1} & 0 & 0 & 0 & 0 & \dots & \dots & G_1 & G_0 \\ \Psi_M & 0 & 0 & 0 & 0 & \dots & \dots & 0 & G_1 \end{bmatrix} \quad (8)$$

$$A_0 = e_M \otimes e'_1 \otimes G_0 \quad (9)$$

$$A_1 = \begin{bmatrix} G_1 & G_0 & 0 & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & G_1 & G_0 & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & 0 & G_1 & G_0 & 0 & \dots & \dots & 0 & 0 \\ 0 & 0 & 0 & G_1 & G_0 & \dots & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \dots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & \dots & G_1 & G_0 \\ G_2 & 0 & 0 & 0 & 0 & \dots & \dots & 0 & G_1 \end{bmatrix} \quad (10)$$

$$A_2 = \begin{bmatrix} G_3 & G_2 & 0 & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & G_3 & G_2 & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & 0 & G_3 & G_2 & 0 & \dots & \dots & 0 & 0 \\ 0 & 0 & 0 & G_3 & G_2 & \dots & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \dots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & \dots & G_3 & G_2 \\ 0 & 0 & 0 & 0 & 0 & \dots & \dots & 0 & G_3 \end{bmatrix} \quad (11)$$

$$G_j = (D_{1-j} \otimes S_1) \dot{+} (D_{1-j} \otimes S_2) \dot{+} \dots \dot{+} (D_{1-j} \otimes S_M); \quad j = 0, 1 \quad (12)$$

$$G_j^* = (D_{3-j} \otimes S_1^0 \beta_M) \dot{+} (D_{3-j} \otimes S_2^0 \beta_M) \dot{+} \dots \dot{+} (D_{3-j} \otimes S_M^0 \beta_M); \quad j = 2, 3 \quad (13)$$

$$G_j = G_j^*(e \otimes e'_M \otimes I); \quad j = 2, 3 \quad (14)$$

$$\Gamma_0 = \begin{bmatrix} D_0 & D_1(1-\theta_1) & 0 & 0 & 0 \dots \dots 0 & 0 \\ 0 & D_0 & D_1(1-\theta_2) & 0 & 0 \dots \dots 0 & 0 \\ 0 & 0 & D_0 & D_1(1-\theta_3) & 0 \dots \dots 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \dots \dots \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \dots \dots \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 \dots \dots D_0 & D_1(1-\theta_{M-1}) \\ 0 & 0 & 0 & 0 & 0 \dots \dots 0 & D_0 \end{bmatrix}$$

$$\Gamma_i = G_4^*(\mathbf{e} \otimes \mathbf{e}'_i \otimes I) + G_{5,i}^*(\mathbf{e} \otimes \mathbf{e}'_{i+1} \otimes I), \quad 1 \leq i \leq M-1 \tag{16}$$

where

$$G_4^* = (D_0 \otimes S_1^0) \dot{+} (D_0 \otimes S_2^0) \dot{+} \dots \dot{+} (D_0 \otimes S_M^0), \text{ and} \tag{17}$$

$$G_{5,i}^* = (1-\theta_i)G_4^*; \quad i = 1, 2, \dots, M \tag{18}$$

$$\Gamma_M = G_4^*(\mathbf{e} \otimes \mathbf{e}'_M \otimes I) \tag{19}$$

$$\Psi_0 = (D_1 \otimes \beta_1 \theta_1) \dot{+} (D_1 \otimes \beta_2 \theta_2) \dot{+} \dots \dot{+} (D_1 \otimes \beta_M \theta_M) \tag{20}$$

$$\Psi_1 = G_{6,1}^*(\mathbf{e} \otimes \mathbf{e}'_1 \otimes I) + G_1 \tag{21}$$

$$\Psi_i = G_{6,1}^*(\mathbf{e} \otimes \mathbf{e}'_i \otimes I); \quad i = 2, 3, \dots, M \tag{22}$$

where

$$G_{6,i}^* = (D_1 \otimes S_1^0 \beta_i \theta_i) \dot{+} (D_1 \otimes S_2^0 \beta_i \theta_i) \dot{+} \dots \dot{+} (D_1 \otimes S_M^0 \beta_i \theta_i) \tag{23}$$

$$i = 1, 2, \dots, M.$$

The matrix  $P$  is of the Quasi-Birth and Death type. The steady state vector  $\mathbf{x}$  partitioned as  $\mathbf{x} = [\mathbf{x}^*, \mathbf{x}(0), \mathbf{x}(1), \mathbf{x}(2), \dots]$  and associated with this matrix is obtained by solving  $\mathbf{x}P = \mathbf{x}$  and  $\mathbf{x}\mathbf{e} = 1$ .

**Theorem:** *The matrix  $P$  is positive recurrent if*

$$\lambda^* < M\mu_M^* \tag{24}$$

**Proof:** The matrix  $A = A_0 + A_1 + A_2$  is block circulant and its invariant probability vector  $\pi$  is simply given as  $\pi = \frac{1}{M}[\phi, \phi, \dots, \phi]$  where  $\phi$  is the invariant probability vector of  $G = G_0 + G_1 + G_2 + G_3$ .

It is know from Neuts [8] that  $\pi A_2 \mathbf{e} > \pi A_0 \mathbf{e}$  if the matrix  $P$  is positive recurrent. Applying this condition, and after routine algebraic manipulations, the stated result follows.

Noting that the matrix  $P$  has a structure that satisfies both the GI/M/1 and the M/G/1 paradigms as described by Neuts [8], we choose the GI/M/1 paradigm to take advantage of the special structure of the matrix  $A_0$ . The vector  $\mathbf{x}(i+1)$  can be determined recursively as  $\mathbf{x}(i+1) = \mathbf{x}(i)R$ , for  $i \geq 0$ , where  $R$  is the minimal non-negative solution to the matrix quadratic

equation:

$$R = A_0 + RA_1 + R^2A_2. \tag{25}$$

The vector  $[\mathbf{x}^*, \mathbf{x}(0)]$  can be obtained as the left eigenvector of the matrix  $B[R]$  where:

$$B[R] = \begin{bmatrix} B_{00} & B_{01} \\ B_{10} & B_{11} + RA_2 \end{bmatrix}. \tag{26}$$

The vector  $[\mathbf{x}^*, \mathbf{x}(0)] = [\mathbf{x}^*, \mathbf{x}(0)]B[R]$  and this vector is normalized by

$$\mathbf{x}^* \mathbf{e} + \mathbf{x}(0)[I - R]^{-1} \mathbf{e} = 1. \tag{27}$$

#### 4. The Matrix $R$

By partitioning  $R$  into blocks of  $nv \times nv$  matrices and noting that the first  $(M - 1)$  blocks of  $A_0$  are zeros,  $R$  can be rewritten as:

$$R = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \dots & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \dots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \dots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & \dots & 0 & 0 \\ R_1 & R_2 & R_3 & R_4 & R_5 & \dots & \dots & R_{M-1} & R_M \end{bmatrix}.$$

Hence,  $R$  can be computed using matrices of smaller dimensions as follows

$$R_1 = G_0 + R_1G_1 + R_MG_2 + R_MR_1G_3, \tag{28}$$

$$R_j = R_{j-1}G_0 + R_jG_1 + R_MR_{j-1}G_2 + R_MR_jG_3, \quad 2 \leq j \leq M \tag{29}$$

**Lemma 1:** We have  $R_j = R_1^j$ ,  $1 \leq j \leq M$ , where  $R_1$  is the solution to

$$R_1 = G_0 + R_1G_1 + R_1^MG_2 + R_1^{M+1}G_3. \tag{30}$$

**Proof:** Equation (28) can be rewritten as:

$$R_1 = (G_0 + R_MR_1G_2 + R_MR_1G_3)(I - G_1)^{-1} \tag{31}$$

where  $I - G_1$  is nonsingular. Replacing  $R_j$  with  $R_1^j$  in (29), it is straightforward to show that  $R_j$  as given by

$$R_j = [(G_0 + R_1^M G_2 + R_1^{M+1} G_3)(I - G_1)^{-1}]^j \quad (32)$$

satisfies Equation (28) and Equation (29). Using the fact that the irreducibility of  $A$  guarantees a unique solution to Equation (28) and Equation (29) (see Neuts [8]), the stated result follows.

Equation (28) can be further simplified by exploiting the structure of  $G_0$ ,  $G_1$ ,  $G_2$  and  $G_3$ . Both  $G_0$  and  $G_1$  are block diagonal matrices and  $G_2$  and  $G_3$  have the first  $(\nu - n_M)$  blocks of columns of zeros. Thus  $R_1$  can be written as

$$R_1 = \begin{bmatrix} R_{11} & 0 & 0 & 0 & 0 & \dots & \dots & 0 & R_{1M} \\ 0 & R_{22} & 0 & 0 & 0 & \dots & \dots & 0 & R_{2M} \\ 0 & 0 & R_{33} & 0 & 0 & \dots & \dots & 0 & R_{3M} \\ 0 & 0 & 0 & R_{44} & 0 & \dots & \dots & 0 & R_{4M} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \dots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & \dots & 0 & R_{MM} \end{bmatrix} \quad (33)$$

where  $R_{jj}$  is of dimension  $(nn_j) \times (nn_j)$  and  $R_{jM}$  is of dimension  $(nn_j) \times (nn_M)$ . The computation of  $R_1$  is carried out as follows. First the diagonal (block) matrices  $R_{jj}$ ,  $1 \leq j \leq M-1$  are computed. Secondly,  $R_{MM}$ , is evaluated and then  $R_{jM}$ ,  $1 \leq j \leq M-1$  is evaluated. The following lemma gives relevant expressions.

**Lemma 2:** *We have*

$$R_{jj} = (D_1 \otimes S_j)[I - D_0 \otimes S_j]^{-1}, \quad 1 \leq j \leq M-1, \quad (34)$$

$$R_{MM} = D_1 \otimes S_M + R_{MM}(D_0 \otimes S_M) + R_{MM}^M(D_1 \otimes S_M^0 \beta_M) + R_{MM}^{M+1}(D_0 \otimes S_M^0 \beta_M), \quad (35)$$

and

$$R_{jM} = R_{jM}(D_0 \otimes S_M) + R_{jj}^M(D_1 \otimes S_j^0 \beta_M) + \sum_{w=1}^M R_{jj}^{M-w} R_{jM} R_{MM}^{w-1}(D_1 \otimes S_j^0 \beta_M) \quad (36)$$

$$+ R_{jj}^{M+1}(D_0 \otimes S_j^0 \beta_M) + \sum_{w=0}^M R_{jj}^{M-w} R_{jM} R_{MM}^w(D_0 \otimes S_j^0 \beta_M), \quad 1 \leq j \leq M-1.$$

**Proof:** Using the form of  $R_1$  as given in Equation (33) in Equation (30), the stated equations follow immediately.

**Special Case:** Consider the case of Bernoulli arrivals with  $D_0 = q_0$  and  $D_1 = p_0$  such that  $p_0 + q_0 = 1$ . Let the service times for a group of size  $i$  be geometric with parameters  $p_i$ ,  $1 \leq i \leq M$ . Then, for this special case we have:

$$G_0 = \Delta(p_0, \mathbf{q}) \quad (37)$$

$$G_1 = \Delta(q_0, \mathbf{q}) \quad (38)$$

$$G_2 = \Delta(p_0, \mathbf{p})(e \otimes e'_M) \quad (39)$$

$$G_3 = \Delta(q_0, \mathbf{p})(e \otimes e'_M) \quad (40)$$

where  $\Delta(k, \mathbf{p}) = \text{diag}[kp_1, kp_2, \dots, kp_M]$ ,  $q_i = 1 - p_i$ ,  $\mathbf{p} = [p_1, p_2, \dots, p_M]$  and  $\mathbf{q} = (q_1, q_2, \dots, q_M)$ .

Hence,

$$R_1 = \begin{bmatrix} r_{11} & 0 & 0 & 0 & 0 & \dots & \dots & 0 & r_{1M} \\ 0 & r_{22} & 0 & 0 & 0 & \dots & \dots & 0 & r_{2M} \\ 0 & 0 & r_{33} & 0 & 0 & \dots & \dots & 0 & r_{3M} \\ 0 & 0 & 0 & r_{44} & 0 & \dots & \dots & 0 & r_{4M} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \dots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & \dots & 0 & r_{MM} \end{bmatrix},$$

where

$$r_{jj} = p_0 q_j (1 - q_0 q_j)^{-1}, \quad 1 \leq j \leq M - 1, \tag{41}$$

$$r_{MM} = p_0 q_M + r_{MM} q_0 q_M + r_{MM}^M p_0 p_M + r_{MM}^{M+1} q_0 p_M, \tag{42}$$

$$r_{jM} = r_{jM} q_0 q_M + r_{jj}^M p_0 p_j + p_0 p_j \sum_{v=1}^M r_{jj}^{M-v} r_{jM}^v r_{MM}^{v-1} + r_{jj}^{M+1} q_0 p_j + q_0 p_j \sum_{v=0}^M r_{jj}^{M-v} r_{jM}^v r_{MM}^v, \quad 1 \leq j \leq M - 1. \tag{43}$$

Note that even for this simple case, there appears to be no explicit expression for  $R_1$ . However, the computation of  $R_1$  is straightforward.

### 5. Performance Measures

1. Using the results from the last section, the vector  $\mathbf{x}(i) = [\mathbf{x}_0(i), \dots, \mathbf{x}_{M-1}(i)]$  can be obtained as

$$\mathbf{x}_j(i+1) = \mathbf{x}_{M-1}(i) R_1^{j+1}, \quad 0 \leq j \leq M - 1, \quad i \geq 1. \tag{44}$$

2. The mean queue length  $\mu_L$  is given as:

$$\mu_L = \mathbf{x}^* (\boldsymbol{\gamma} \otimes \mathbf{e}) + M \mathbf{x}(0) R (I - R)^{-2} \mathbf{e} + \mathbf{x}(0) (I - R)^{-1} (\boldsymbol{\gamma} \otimes \mathbf{e}) \tag{45}$$

where

$$\boldsymbol{\gamma} = [0, 1, 2, 3, \dots, M - 1]' \tag{46}$$

and

$$\mathbf{x}^* = [\mathbf{x}_0^*, \mathbf{x}_1^*, \mathbf{x}_2^*, \dots, \mathbf{x}_{M-1}^*]. \tag{47}$$

Note that  $(I - R)^{-1}$  is obtained as

$$(I - R)^{-1} = \begin{bmatrix} I & 0 & \dots & 0 & 0 \\ 0 & I & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & I & 0 \\ (I - R_M)^{-1}R_1 & (I - R_M)^{-1}R_2 & \dots & (I - R_M)^{-1}R_{M-1} & (I - R_M)^{-1} \end{bmatrix}$$

3. The probability that the server is on vacation (or idle),  $y_0$  is given by  $y_0 = \mathbf{x}^* \mathbf{e}$ .

4. The probability that the number of customers waiting is less than  $kM - 1$ ,  $y_k$ , is given by  $y_k = \sum_{i=1}^k \mathbf{x}(i-1)\mathbf{e}$ .

### 6. Numerical Examples

#### 6.1 PH/Geo/1 System

In this example we consider a PH/Geo/1 system, and our interest is to determine the effects that various theta values would have on the system performance measures.

Four different sets of values for theta were considered, 1) 0/1 type, 2) linear, 3) convex, 4) concave (see Appendix A). In all cases, we take  $\theta_j \leq \theta_{j+1}$ ;  $j = 0, 1, \dots, M - 1$ .

Define

$$\theta^i(v) = [\theta_1^i(v)\theta_2^i(v)\dots\theta_{M-1}^i(v)] \tag{48}$$

- $v = 1$  for 0/1 type
- $v = 2$  for linear
- $v = 3$  for convex
- $v = 4$  for concave.

In all cases we take

$$\theta_j^i(v) = 0, j < i, 0 < \theta_j^i(v) \leq \theta_{j+1}^i(v) \leq 1, j \geq i, \tag{49}$$

where  $i \geq 1$  is a lower threshold before service is even considered.

**Example 1:** In this example we take  $M = 8$ ,  $D_0$  and  $D_1$  are as follows,

$$\mathbf{C}_0 = \begin{bmatrix} 0.2 & 0.6 \\ 0.3 & 0.4 \end{bmatrix} \quad \mathbf{C}_1 = \begin{bmatrix} 0.06 & 0.14 \\ 0.09 & 0.21 \end{bmatrix}.$$

The vector of the service rate is [0.66, 0.4356, 0.2875, 0.1897, 0.1252, 0.0827, 0.0546, 0.036]. The arrival rate for this is 0.2655. Figures 1 and 2 display the two performance measures, mean queue length and server idle probability as a function of the group size. An examination of these figures

reveal the following observations. For high traffic where  $\mu_M = 0.036$ , the shortest queue length is for  $\theta^i(1)$ . Also  $y_0$  is smallest when group size is  $M$ .

**Example 2:** In this example we consider a system with low traffic by taking  $\mu_M = 0.101$ . Arrivals have the same MAP representation as in Example 1 and the vector of the service rate is  $[0.75, 0.5625, 0.4219, 0.3164, 0.2373, 0.1778, 0.1335, 0.101]$ . Figures 3 and 4 display the two performance measures, mean queue length and server idle probability, as a function of group size. An examination of these figures reveal the following observations. For high traffic where  $\mu_M = 0.101$  the shortest queue length is for  $\theta^i(1)$ . Also  $y_0$  is smallest when group size is  $M$ .

Figures 1 through 4 show that the best performance was found when theta values are 0/1. The results indicate that a shorter queue length results when there are specific guidelines that tell the server when to serve the customers.

## 6.2 MAP/PH/1 System

Since we determined from the PH/Geo/1 examples that theta values of 0/1 type result in the shortest queue values, we will only consider 0/1 type values for the thetas in the remaining examples (see Appendix B).

$$\theta^i = \theta^i(1) = [\theta_1^i, \theta_2^i, \dots, \theta_1^i]. \tag{50}$$

We take  $\theta_j^i = 0, j < i$  and  $\theta_j^i = 1, j \geq i$ .

For example, if  $j = 3$ , the server cannot start serving unless there are at least 3 customers waiting for service.

### 6.2.1 Arrival Rates

1.  $\lambda^* = 0.9$

$$D_0 = \begin{bmatrix} 0.1 & 0.0 \\ 0.05 & 0.05 \end{bmatrix} \quad D_1 = \begin{bmatrix} 0.8 & 0.1 \\ 0.3 & 0.6 \end{bmatrix}.$$

2.  $\lambda^* = 0.7$

$$D_0 = \begin{bmatrix} 0.3 & 0.1 \\ 0.1 & 0.1 \end{bmatrix} \quad D_1 = \begin{bmatrix} 0.4 & 0.2 \\ 0.3 & 0.5 \end{bmatrix}.$$

3.  $\lambda^* = 0.5$

$$D_0 = \begin{bmatrix} 0.3 & 0.3 \\ 0.1 & 0.3 \end{bmatrix} \quad D_1 = \begin{bmatrix} 0.3 & 0.1 \\ 0.2 & 0.4 \end{bmatrix}.$$

4.  $\lambda^* = 0.2$

$$D_0 = \begin{bmatrix} 0.6 & 0.2 \\ 0.3 & 0.3 \end{bmatrix} \quad D_1 = \begin{bmatrix} 0.1 & 0.1 \\ 0.3 & 0.1 \end{bmatrix}.$$

All the following examples will use the above arrival representations. Service times have a

Negative Binomial distribution which is a special case of a discrete phase distribution and will be indicated in the examples. The Negative Binomial distribution is the discrete analog of the Erlang distribution. As group size increases so does the average service time. Our selection of the Negative Binomial distribution for service is due to its more common occurrence in real life. This is not a limitation of our model and our model allows a general phase type distribution.

**Example 3:** In this example we take phase service to be of the same dimension  $n_i = 2$  phases, ( $i = 1, \dots, 10$ ).  $M = 10, \beta = \beta_i = [1, 0]$  ( $i = 1, \dots, 10$ ), and theta values are of the form 0/1 type. Our interest in this example is to study the system for the case where phase service has the same dimension for the different group sizes even though the mean service times are increasing with group size.

Table 2 shows the service times for each group size  $i$ ;  $\mu_{S_i}^{-1}$  is the mean service time and  $\sigma_{S_i}^2$  is the variance of the service time.

The results are shown in Figure 5. The mean queue length increased as the group size increased. According to this example, the smallest mean number in the queue resulted when the group size was 1. The smallest  $y_0$  value (probability the server is idle) also resulted when the group size was 1. As the traffic intensity increased (from 0.2 to 0.9) so did the mean number in the queue and  $y_0$  (see Figure 6).  $y_0$  increases initially as  $i$  increases and then it decreases as  $i$  increases. In this example, both the mean and variance of the service time increased monotonically with  $i$ .

**Table 2 - Service times for group size  $i$**

$i$	$S_i$	$\mu_{S_i}$	$\sigma_{S_i}^2$
1	$\begin{bmatrix} 0.2 & 0.8 \\ 0.0 & 0.2 \end{bmatrix}$	0.4	0.625
2	$\begin{bmatrix} 0.27 & 0.73 \\ 0.0 & 0.27 \end{bmatrix}$	0.365	1.013
3	$\begin{bmatrix} 0.34 & 0.66 \\ 0.0 & 0.34 \end{bmatrix}$	0.33	1.561
4	$\begin{bmatrix} 0.4 & 0.6 \\ 0.0 & 0.4 \end{bmatrix}$	0.3	2.222
5	$\begin{bmatrix} 0.47 & 0.53 \\ 0.0 & 0.47 \end{bmatrix}$	0.265	3.346
6	$\begin{bmatrix} 0.54 & 0.46 \\ 0.0 & 0.54 \end{bmatrix}$	0.23	5.104
7	$\begin{bmatrix} 0.6 & 0.4 \\ 0.0 & 0.6 \end{bmatrix}$	0.2	7.5
8	$\begin{bmatrix} 0.67 & 0.33 \\ 0.0 & 0.67 \end{bmatrix}$	0.165	12.305

9	$\begin{bmatrix} 0.74 & 0.26 \\ 0.0 & 0.74 \end{bmatrix}$	0.13	21.893
10	$\begin{bmatrix} 0.8 & 0.2 \\ 0.0 & 0.8 \end{bmatrix}$	0.1	40.0

**Example 4:** In this example, we allow the dimension of the phase service to vary with the group size and take  $n_i = 1, i = 1, 2, 3; n_i = 2, i = 4, 5, 6, 7; n_i = 3, i = 8, 9, 10$ .  $M = 10, \beta_i = 1, i = 1, 2, 3; \beta_i = [0, 1], i = 4, 5, 6, 7; \beta_i = [1, 0, 0], i = 8, 9, 10$ , and theta values chosen as 0/1 type. In this example, the mean of the service time for each group size  $i$  is the same as that for the corresponding group size in Example 3. However, the variance of the service time is non-monotonic.

Table 3 shows the service times for each group size  $i$  and the corresponding mean and variance.

**Table 3 - Service times for group size  $i$**

$i$	$S_i$	$\mu_{S_i}$	$\sigma_{S_i}^2$
1	0.6	0.4	3.75
2	0.635	0.365	4.766
3	0.667	0.333	6.015
4	$\begin{bmatrix} 0.4 & 0.6 \\ 0.0 & 0.4 \end{bmatrix}$	0.3	2.222
5	$\begin{bmatrix} 0.47 & 0.53 \\ 0.0 & 0.47 \end{bmatrix}$	0.265	3.346
6	$\begin{bmatrix} 0.54 & 0.46 \\ 0.0 & 0.54 \end{bmatrix}$	0.23	5.104
7	$\begin{bmatrix} 0.6 & 0.4 \\ 0.0 & 0.6 \end{bmatrix}$	0.2	7.5
8	$\begin{bmatrix} 0.505 & 0.495 & 0.0 \\ 0.0 & 0.505 & 0.495 \\ 0.0 & 0.0 & 0.505 \end{bmatrix}$	0.165	6.183
9	$\begin{bmatrix} 0.61 & 0.39 & 0.0 \\ 0.0 & 0.61 & 0.39 \\ 0.0 & 0.0 & 0.61 \end{bmatrix}$	0.13	12.032

10	$\begin{bmatrix} 0.7 & 0.3 & 0.0 \\ 0.0 & 0.7 & 0.3 \\ 0.0 & 0.0 & 0.7 \end{bmatrix}$	0.1	23.333
----	---	-----	--------

The results are shown in Figure 7. According to this example, the minimum mean number in the queue resulted when the group size was 1 for low to medium-high traffic. However, for high (0.9) traffic intensity, the minimum mean number in the queue resulted when the group size was 4. Note that the service time of group size 4 has the smallest variance. This seems to have a major impact on the performance measures at a high traffic level. Increasing the mean service time did not necessarily result in an increase in mean queue length, the mean queue length either decreases or did not increase at an expected rate when the variance of the service time is decreased. This is to be expected. As shown in Figure 8, in all cases server idleness probability was smallest with a group size of 1, with the lowest value obtained with a traffic intensity of 0.9.

## 7. The Stationary Waiting Time Distribution

In this section we present the equations for obtaining the waiting time distribution. The computational aspect is intensive and is a subject of future work.

Before obtaining the waiting time distribution we need to set up some notations. Let  $\mathbf{y}_j^*$  denote the probability vector that an arriving customer finds the server idle with  $j$  waiting in the queue. Similarly  $\mathbf{y}_{i,j'}(i)$ ,  $i \geq 0$ ,  $0 \leq j \leq M-1$ ,  $1 \leq j' \leq M$  is defined as the probability vector that an arriving customer finds the server busy with  $j'$  customers and there are  $iM + j$  customers waiting. Then we have

$$\mathbf{y}^* = \frac{1}{\lambda^*} \mathbf{x}(0)(I \otimes D_1), \quad i \geq 0, \quad (51)$$

$$\mathbf{y}(i) = \frac{1}{\lambda^*} \mathbf{x}(i)(I \otimes D_1), \quad i \geq 0, \quad (52)$$

where

$$\begin{aligned} \mathbf{y}^* &= [\mathbf{y}_0^*, \mathbf{y}_1^*, \dots, \mathbf{y}_{M-1}^*], \\ \mathbf{y}_j(i) &= [\mathbf{y}_{j,1}(i), \mathbf{y}_{j,2}(i), \dots, \mathbf{y}_{j,M}(i)], \\ \mathbf{y}(i) &= [\mathbf{y}_0(i), \mathbf{y}_1(i), \dots, \mathbf{y}_{M-1}(i)], \end{aligned}$$

and  $\lambda^*$  is the arrival rate.

Further, let  $P(r, k)$  denote the matrix of dimension  $n \times n$  whose  $(i, j)^{th}$  entry is given by  $P_{ij}(r, k) = Pr\{N(k) = r, J_k = j \mid N(0) = 0, J_0 = i\}$ , where  $N(k)$  = number of arrivals during  $(0, k]$  and  $J_k$  = phase of the arrival process at time  $k$ . It is easy to verify that

$$P(r, k) = P(r, k-1)D_0 + P(r-1, k-1)D_1, \quad r \geq k+1 \quad (53)$$

where

$$P(0, r) = \begin{cases} I, & r = 0 \\ 0, & r \neq 0 \end{cases}$$

$$P(k, k) = D_1^k, \quad k \geq 0$$

$$P(0, k) = D_0^k, \quad k \geq 0.$$

Now let  $w(k)$  denote the probability mass function of the waiting time of a customer at an arrival.

**Theorem 2:** *The pmf  $w(i)$  of the waiting time of a customer at an arrival is given by*

$$w(k) = \begin{cases} \sum_{j=0}^{M-1} \mathbf{y}_j^* \mathbf{e} \theta_{j+1}, & k = 0 \\ w_{11}(k), & k = 1 \\ w_{11}(k) + w_{21}(k) + w_{31}(k), & k = 2 \\ w_{11}(k) + w_{22}(k) + w_{31}(k) + w_{32}(k), & k \geq 3 \end{cases}$$

where

$$w_{11}(k) = w_1(k) + w_{21}(k) \tag{54}$$

with

$$w_1(k) = \sum_{j=0}^{M-2} \sum_{r=0}^{M-2-j} \mathbf{y}_j^* P(r, k-1) D_1 \mathbf{e} \hat{\theta}_{r+j+2}, \quad k \geq 1, \tag{55}$$

where

$$\hat{\theta}_{r+j} = (1 - \theta_{j+1})(1 - \theta_{j+2}) \dots (1 - \theta_{r+j-1}) \theta_{r+j},$$

$$w_{21}(k) = \sum_{j=0}^{M-2} \sum_{j_1=1}^M \mathbf{y}_{j, j_1}(0) \sum_{r=0}^{M-2-j} (P(r, k-1) D_1 \mathbf{e} \otimes S_{j_1}^{k-1} \mathbf{S}_{j_1}^0) \theta_{r+1+j} \tag{56}$$

$$+ \sum_{j=0}^{M-1} \sum_{j_1=1}^M \mathbf{y}_{j, j_1}(0) \sum_{r=M-1-j}^{\infty} (P(r, k) \mathbf{e} \otimes S_{j_1}^{k-1} \mathbf{S}_{j_1}^0), \quad k \geq 1,$$

$$w_{22}(k) = \sum_{j=0}^{M-2} \sum_{j_1=1}^M \mathbf{y}_{j, j_1}(0) \sum_{u=0}^{k-1} \sum_{r=0}^{M-2-j} (P(r, u) \otimes S_{j_1}^{u-1} \mathbf{S}_{j_1}^0) \tag{57}$$

$$\times \sum_{s=0}^{M-2-r-j} P(s, k-1-u) D_1 \mathbf{e} \hat{\theta}_{j+r+s+2}, \quad k \geq 2,$$

$$w_{31}(k) = \sum_{j=0}^{M-2} \sum_{j_1=1}^M \sum_{i=1}^{k-1} \sum_{u=1}^{k-i} \mathbf{y}_{j, j_1}(i)$$

$$\times \sum_{r=0}^{M-2-j} (P(r, k-1) D_1 \mathbf{e} \otimes S_{j_1}^{u-1} \mathbf{S}_{j_1}^0) \delta(i) T^{k-1-u}(i) \mathbf{T}^0(i) \theta_{r+j+1} \tag{58}$$

$$+ \sum_{j=0}^{M-1} \sum_{j_1=1}^M \sum_{i=1}^{k-1} \sum_{u=1}^{k-i} \mathbf{y}_{j, j_1}(i) \sum_{r=M-1-j}^{\infty} (P(r, k) \mathbf{e} \otimes S_{j_1}^{u-1} \mathbf{S}_{j_1}^0) \delta(i) T^{k-1-u}(i) \mathbf{T}^0(i), \quad k \geq 2,$$

where  $(\delta(i), T(i))$  of order  $n_M i$  is the representation of the convolution of  $i$  discrete PH-distributions, all with the same representation  $(\beta_M, S_M)$ , and

$$w_{32}(k) = \sum_{j=0}^{M-2} \sum_{j_1=1}^M \sum_{i_1=1}^{k-2} \sum_{u_2=i_1+1}^{k-1} \sum_{u_1=1}^{u_2-i} \sum_{r=0}^{M-2-j} \mathbf{y}_{j, j_1}(i) (P(r, u_2) \otimes S_{j_1}^{u_1-i} S_{j_1}^0) \quad (59)$$

$$\times \delta(i) T^{u_2-u_1-1}(i) T^0(i) \sum_{s=0}^{M-2-r-j} P(s, k-1-u_2) D_1 e^{\hat{\theta}_{r+j+s+2}}, \quad k \geq 3.$$

**Proof:** Follows immediately on noting the following.

1. For  $w(0)$ , i.e., the customer gets into service immediately upon arrival: In this case the server is idle when the customer arrives and there are  $j$ ,  $0 \leq j \leq M-1$  customers waiting in the queue. The server therefore starts to serve  $j+1$  customers with probability  $\theta_{j+1}$ .

2. For  $w_1(k)$ ,  $k \geq 1$ : the customer arrives to find the server idle with  $j$ ,  $0 \leq j \leq M-2$  customers waiting. The customer waits and enters service after  $k$  units such that during the first  $k-1$  units of time  $r$ ,  $0 \leq r \leq M-2-j$ , customers arrive, and at time  $k$  an arrival triggers the service.

3. For both  $w_{21}(k)$ ,  $k \geq 1$  and  $w_{22}(k)$ ,  $k \geq 2$ : the customer arrives to find the server busy with  $j_1$  customers and there are  $j$  customers waiting in the queue.

(a) For  $w_{21}(k)$ ,  $k \geq 1$ : there are two mutually exclusive events to consider.

i) The current service lasts exactly  $k$  units, and  $0 \leq j \leq M-2$ . The customer enters service  $k$  units after arrival with another customer who arrives at time  $k$ . During the first  $k-1$  units of wait  $r$ ,  $0 \leq r \leq M-2-j$ , customers arrive.

ii) Here  $j$ ,  $0 \leq j \leq M-1$ , customers are waiting in the queue. The current service lasts  $k$  units. During the first  $k$  units at least  $M-j$  customers arrive.

(b) For  $w_{22}(k)$ ,  $k \geq 2$  the server ends service at time  $u$ ,  $1 \leq u \leq k-1$ . The server is idle between time  $u$  and  $k$  and initiates a service at time  $k$  with an arrival. During the interval  $(0, u]$  there are  $r$ ,  $0 \leq r \leq M-2-j$ , arrivals and during the interval  $(u, k-1]$  there are  $s$ ,  $0 \leq s \leq M-2-r-j$ , arrivals.

4. For both  $w_{31}(k)$ ,  $k \geq 2$  and  $w_{32}(k)$ ,  $k \geq 3$ : the customer arrives to find the server busy with  $j_1$  customers and there are  $iM+j$ ,  $i \geq 1$ , customers waiting in the queue.

(a) For  $w_{31}(k)$ ,  $k \geq 2$ : the current service lasts  $u$  units and the service time of the  $i$  groups of size  $M$  lasts  $k-u$  units. The customer enters service  $k$  units after arrival at which time another customer arrives. We have to consider two mutually exclusive events.

i) During the first  $k-1$  units  $r$ ,  $0 \leq r \leq M-2-j$ , customers arrive,  $0 \leq j \leq M-2$ ,

ii) During the  $k$  units wait at least  $M-1-j$  customers arrive,  $0 \leq j \leq M-1$ .

(b) For  $w_{32}(k)$ ,  $k \geq 3$ : the server ends the current service at time  $u_1$  and serves the  $i$  groups of  $M$  customers in  $u_2-u_1$ ,  $1 \leq u_1 \leq u_2-i$ ,  $i+1 \leq u_2 \leq k-1$  units of time. The customer enters service at time  $k$  after arrival with another customer who arrives at time  $k$ . During the first  $u_2$  units  $r$ ,  $0 \leq r \leq M-2-j$ , customers arrive and during the next  $k-1-u_2$  units  $s$ ,  $0 \leq s \leq M-2-r-j$ , customers arrive.



## 3. Convex

$$\theta^{i(3)} = \begin{cases} 0.5 & 0.505 & 0.51 & 0.53 & 0.56 & 0.64 & 0.8 & 1.0 & i=1 \\ 0.0 & 0.5 & 0.505 & 0.51 & 0.52 & 0.60 & 0.76 & 1.0 & i=2 \\ 0.0 & 0.0 & 0.5 & 0.505 & 0.525 & 0.59 & 0.74 & 1.0 & i=3 \\ 0.0 & 0.0 & 0.0 & 0.5 & 0.51 & 0.56 & 0.69 & 1.0 & i=4 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.5 & 0.53 & 0.64 & 1.0 & i=5 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.5 & 0.58 & 1.0 & i=6 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.5 & 1.0 & i=7 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & i=8 \end{cases}$$

## 4. Concave

$$\theta^{i(4)} = \begin{cases} 0.5 & 0.75 & 0.84 & 0.89 & 0.93 & 0.96 & 0.98 & 1.0 & i=1 \\ 0.0 & 0.5 & 0.77 & 0.86 & 0.91 & 0.95 & 0.98 & 1.0 & i=2 \\ 0.0 & 0.0 & 0.5 & 0.81 & 0.89 & 0.94 & 0.95 & 1.0 & i=3 \\ 0.0 & 0.0 & 0.0 & 0.5 & 0.84 & 0.92 & 0.94 & 1.0 & i=4 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.5 & 0.93 & 0.94 & 1.0 & i=5 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.5 & 0.93 & 1.0 & i=6 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.5 & 1.0 & i=7 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & i=8 \end{cases}$$

**B. Theta Values for MAP/PH/1 System**

## 1. 0/1 type

$$\theta^{i(1)} = \begin{cases} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & i=1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & i=2 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & i=3 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & i=4 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & i=5 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & i=6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & i=7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & i=8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & i=9 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & i=10 \end{cases}$$

Figure 1 - Queue Length for PH/Geo/1

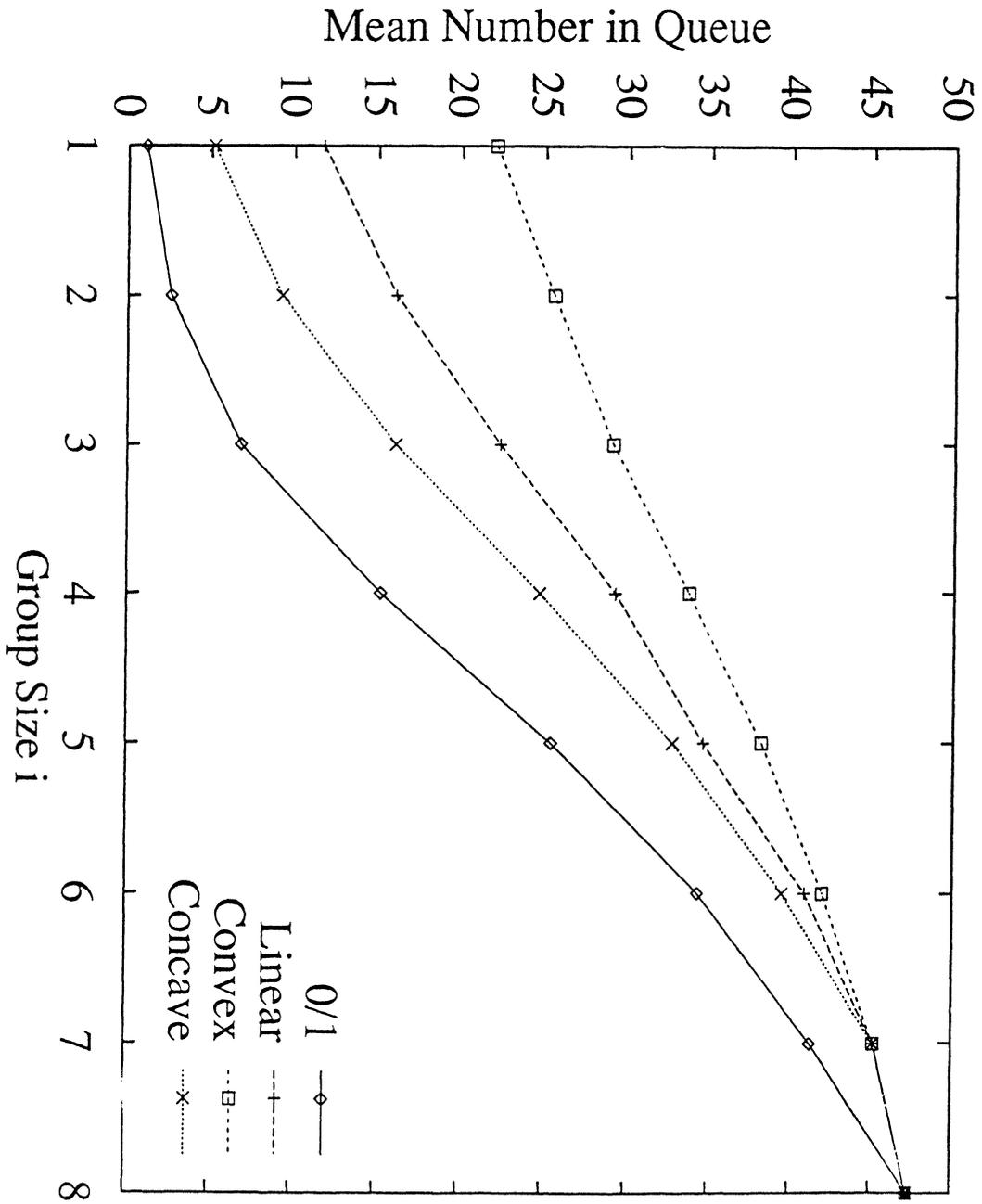


Figure 2 - Probability for Idleness of Server

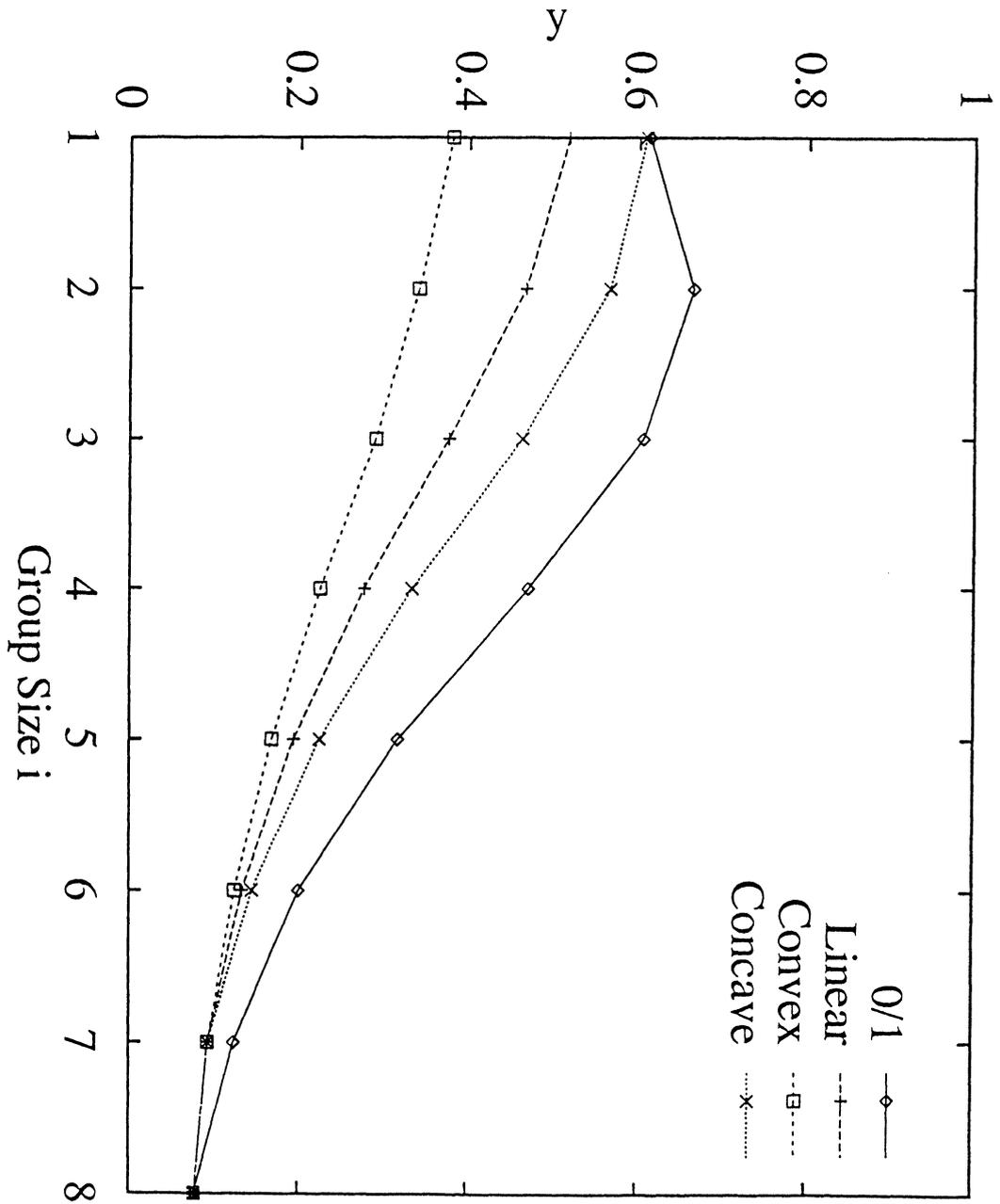


Figure 3 - Queue Length for PH/Geo/1

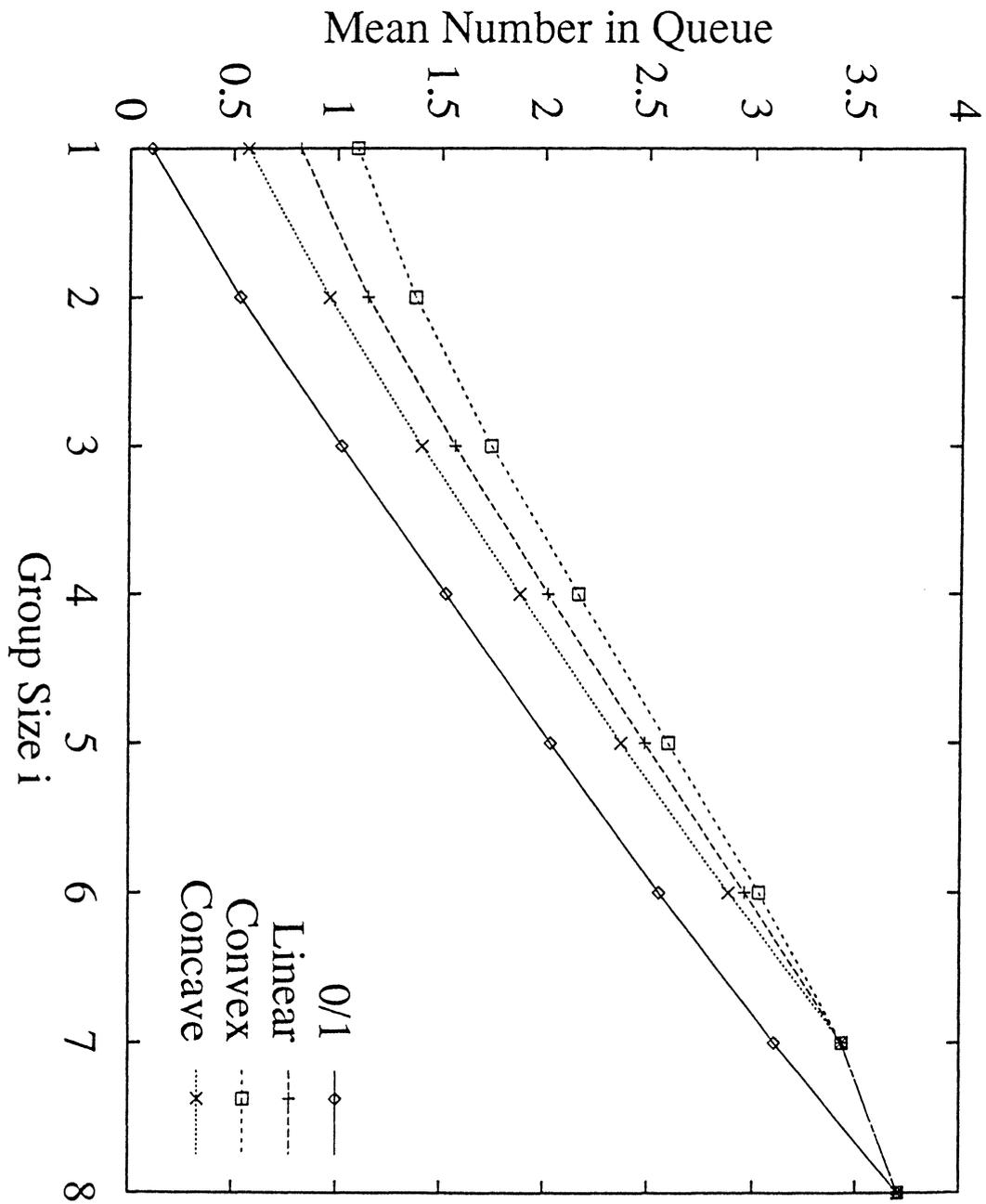


Figure 4 - Probability for Idleness of Server

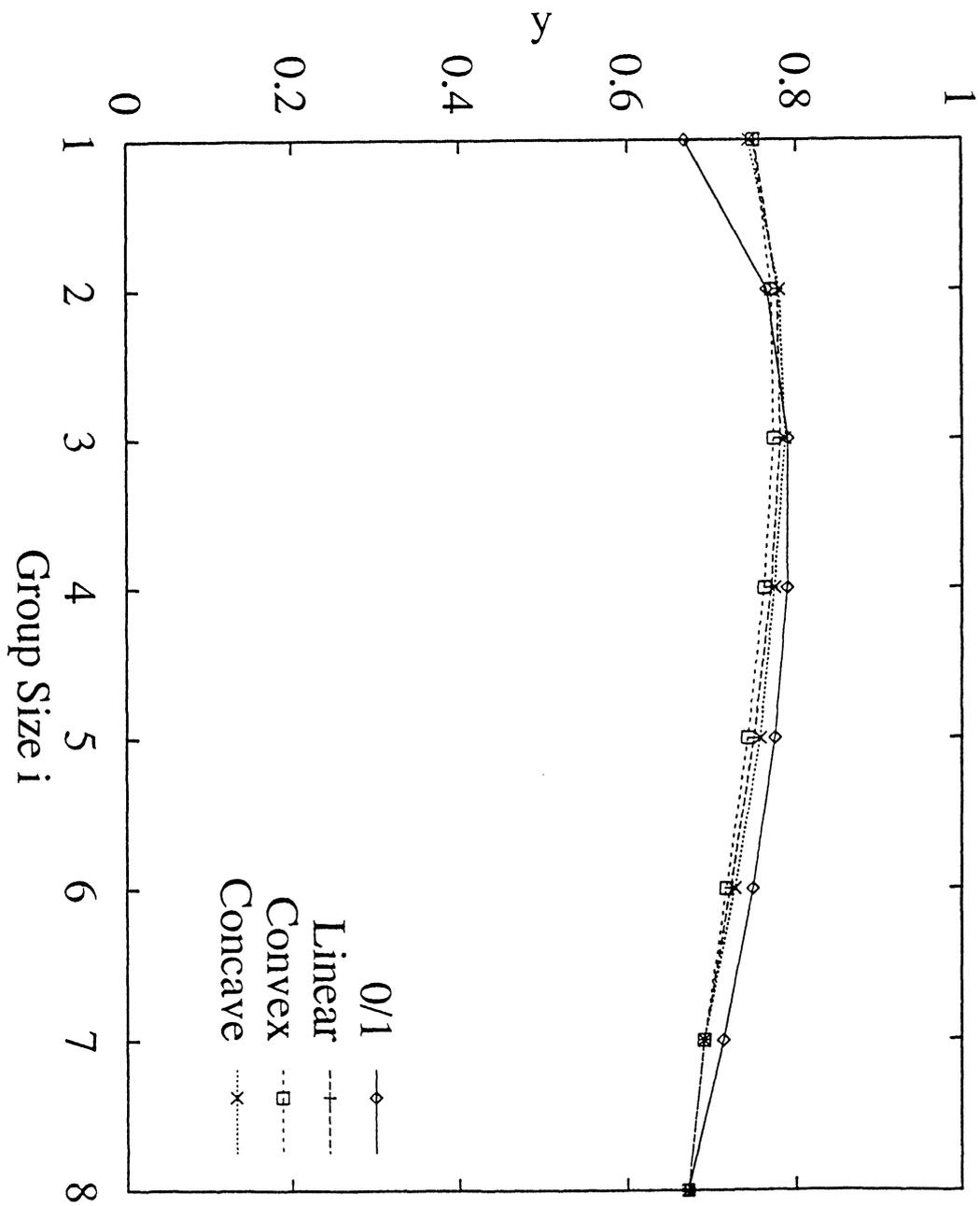


Figure 5 - Queue Length for Phase Type Distribution of Same Size

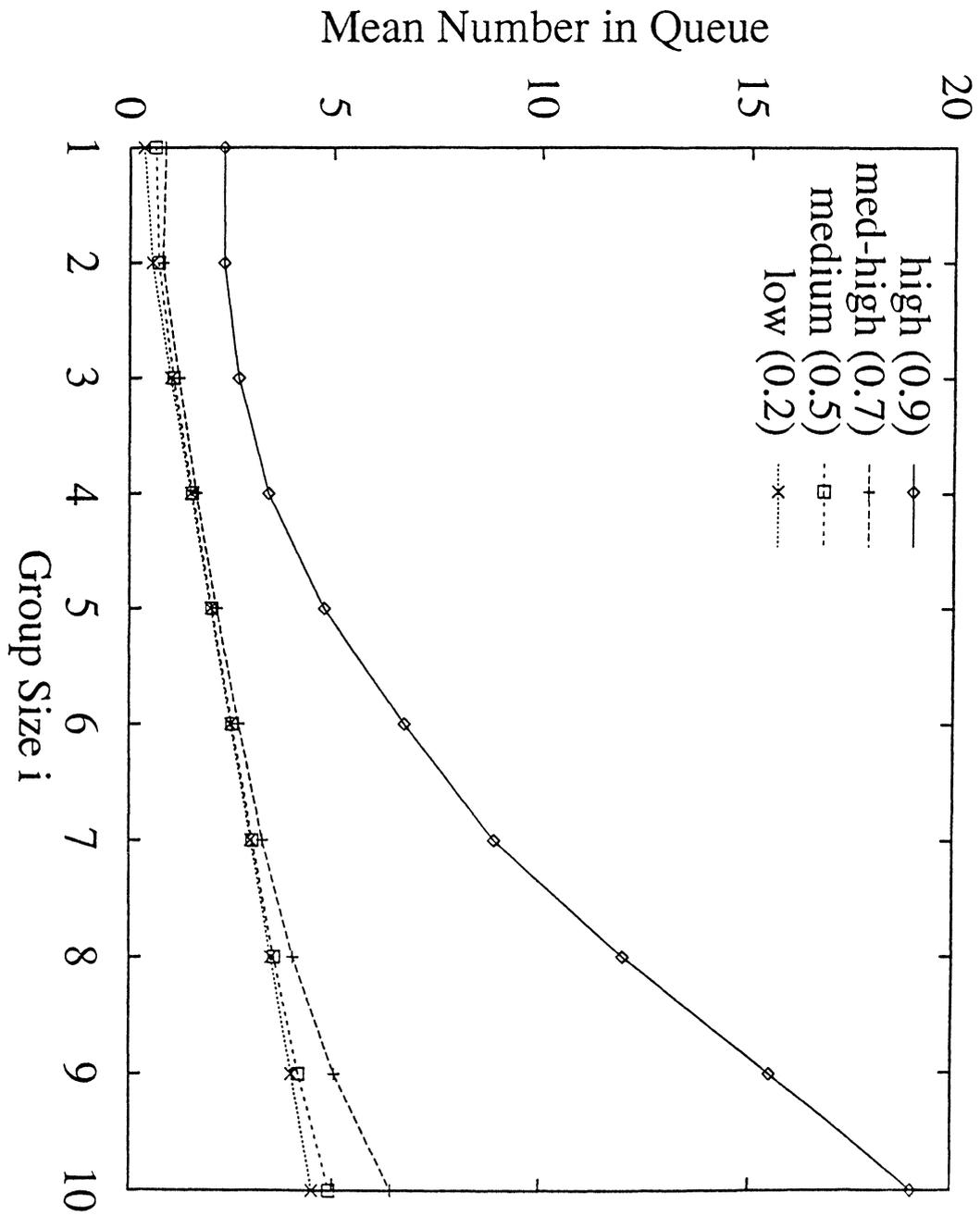


Figure 6 - Probability for Idleness of Server

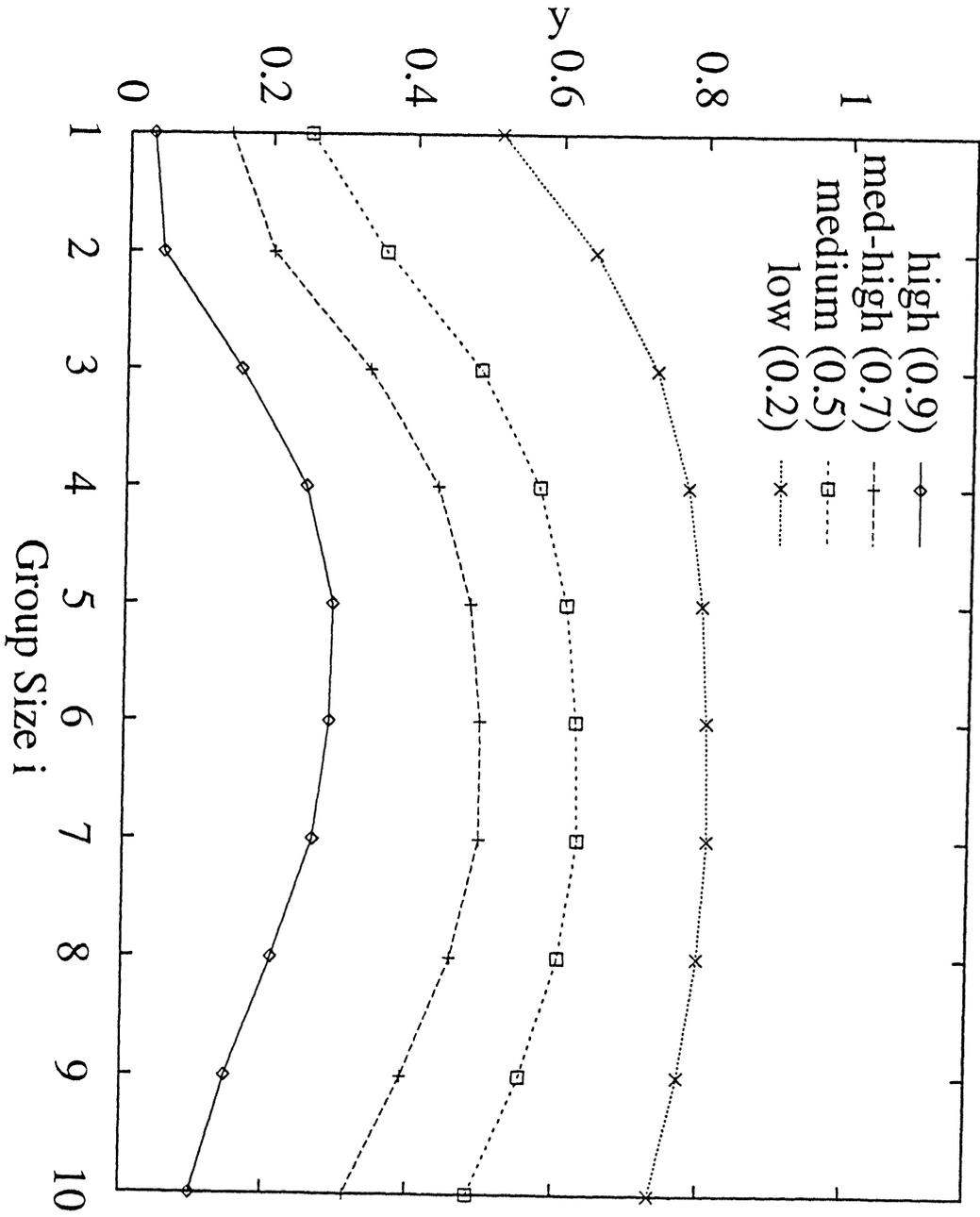


Figure 7 - Queue Length for Phase Type Distribution of Varying Size

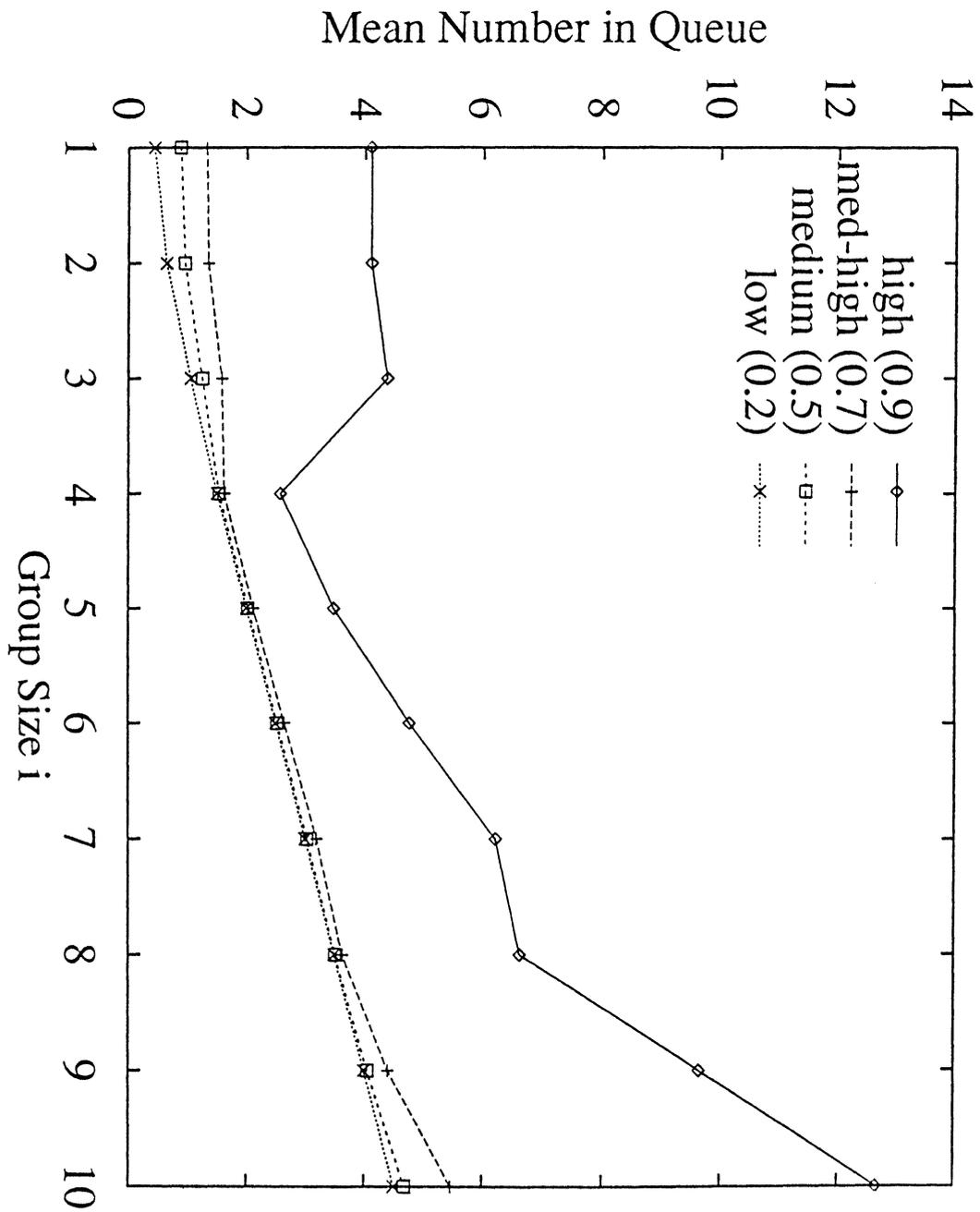


Figure 8 - Probability for Idleness of Server

