Studies on Queueing Systems with Correlated Service Times and Several Types of Customers

相関のあるサービス時間と複数種類のカスタマーの待ち行列システムに関する研究

1996

Kanji Yoneyama

米山　寛二

Department of Applied Physics, Osaka University

Preface

The systematic study of queueing theory was on the design of automatic telephone exchanges at the beginning of this century. It is Erlang [6] who may be considered as the founder of queueing theory. Queueing theory then developed extensively with the progress of operations research. Various queueing systems appeared in technology and management, such as production lines, transportation systems, data processing, computer systems and telecommunication systems, are recognized as a new fields of its application. A vast amount of literature on queueing theory is growing rapidly as the fields of application expand.

In most of the published studies of queueing systems, it has been assumed that the service times of all servers are mutually independent. In some practical situation, however, this assumption is not realistic, because of the competition or cooperation among servers . In this sense, it has been expected to develop the analysis of queueing systems with correlated service times. On the other hand, when describing a queueing system, the concept of customer types is very useful. For instance, most retrial queueing models deal with one type of calls. But there are some practical models which deal with several types of calls. Roughly speaking, this thesis provides explicit solutions for these two problems.

The main theme of this thesis is twofold: first, to introduce the multivariate exponential distribution of Marshall and Olkin [58] as the correlated service time distribution of multiserver and tandem queueing systems and investigate the effect of correlated service times comparing with the independent case; second, to propose several types of tandem queueing systems with the number of customers and study the sensitivity of performance measure for these systems. Some kind of matters which are concerned with queueing theory, such as jockeying, blocking and switching rule etc., are also considered to advance the above work. A multiserver queueing system with additional service channels added to the thesis is such a topic, where the number of channels depends on the present number of customers. The author hopes the works contained in this thesis will contribute to the development of analytical approaches for queueing systems with correlated service times and several types of customers, and be used for practical applications.

Kanji YONEYAMA

December, 1996

Acknowledgement

The author would like to express his sincere appreciation to Professor H. Ishii for supervising this thesis. His continuous encouragement and invaluable comments have helped to accomplish this thesis.

The author also wishes to acknowledge Professor K. Kinoshita and Professor A. Yagi for their useful advice and helpful comments for improving the thesis.

The author is highly indebted to Professor Emeritus T. Nishida of Osaka University, Professor at Osaka International University, who guided the author to the present study and giving the continuous and invaluable suggestions.

Furthermore, the author would like to express his heartfelt gratitude to President S. Motokura and the author's colleague of Himeji College of Hyogo, who gave the author their continuous encouragement.

Finally, the author wishes to thank Associate Professor S. Shiode and the member of Ishii's Laboratory of Osaka University for their continuous kindness and friendship.

Contents

Chapter 1.

INTRODUCTION

1.1. Review of Correlated Queueing Systems

With few exceptions, queueing theory has progressed under three sets of independence assumptions; an arrival process is a renewal process, a service time process is a sequence of i.i.d.(i.e. independent and identically distributed) random variables, arrival processes and service time processes are independent processes.

First, we review the literature on queueing systems which relaxes the assumption of an independent interarrival time.

It is not uncommon, in computer system modelling, to assume that the arrivals of jobs follow a Poisson process. Although this hypothesis is usually chosen because of the resulting tractability, it is recognized that the Poisson process is often a realistic representation of actual arrival process.

Mathematically, this can be explained as follows. In some circumstances, the jobs arrival process actually results from superposition of several independent process. It is well known that the superposition of n renewal processes yields a new process, which is asymptotically a Poisson process, as n increases to infinity, provided that suitable conditions are satisfied (Cox and Smith [15]). If this limiting property does not hold, one concludes firstly that the combined process is not necessarily a Poisson process; then it is not solely characterized by the mean interarrival time. Secondly, the combined process is not necessarily a renewal process, in which case the interarrival times are not independent identically distributed random variables.

Latouche [55] considered a simple queueing system for which the arrival process departs slightly from the Poisson process; the standard deviation of the interarrival time distribution is approximately equal to its mean, and successive interarrival times are almost non-correlated. He showed that the system has a stationary probability vector of matrix-geometric form.

Latouche [56] considered the $M/PH/1$ queueing system whose arrival process is the same as the above. He showed how the stationary probability distribution may be studied by a perturbation analysis.

Szekli et al. [98] considered Markov renewal arrival processes with particular transition matrix for the underlying Markov chain which allow to change dependency properties without changing distribution conditions at the same time. He studied the effects of dependency (such as association) in the arrival process to a single server queueing system on the mean queue length and the mean waiting times.

Daganzo and Diez-Roux [16] examined the infinite-server queueing model with a stationary but arbitrarily correlated arrival process. Their research was motivated by a container storage problem at seaports. They assumed that the arrival process is a point process. They presented exact expressions for the variance of the number of customers in the system.

Next, we survey the study of queueing systems with correlated service times.

As far as we know, Mitchell et al. [60] were the first to investigate the effect of correlation between service times at successive stages on the physical characteristics of tandem queueing systems. That correlation is intuitively clear from the supermarket example where most customers who spend a long time in the shopping area also ipso facto require a long check time. And a patient with unknown disease may spend a long time at each of a series of investigative stages until the disease is identified, followed perhaps by a long period of treatment and convalescence. Mitchell et al. [60] simulated a correlated two-stage tandem queueing model in which their service distribution is a special case of the bivariate exponential distribution discussed by Wicksell [101] and Kibble [45] and more generally by Krishnamoorthy and Parthasarty [50] and Paulson [85]. Their simulation indicates that the expected waiting time at the second stage with infinite intermediate buffers is smaller than in the case of independent service times.

Choo and Conolly [11] studied analytically the above model and showed that the mean waiting time in queue decreases as correlation between the service times increases, for light as well as heavy traffic.

These papers, and development of large-scale data transmission systems, renewed interest in obtaining analytical results for this problem.

The transmission of messages over a network was modeled by Kleinrock [46] as a network queue, where the transmission time of a message at a station is the service time there. Kleinrock observed that the service times of the same customer at different stations are dependent because the transmission time of a message is roughly proportional to message length.

Pinedo and Wolff [87] considered a two-stage tandem queueing system in which a given customer has equal service times at each station. They established that for heavy traffic, the mean waiting time in queue is lower when service times are dependent (equal) than when they are independent, but for light traffic, the reverse is true. They showed that the crossover between these two modes of behavior occurs at server utilization of approximately 0.58.

Wolff [102] obtained light traffic results about the mean waiting time in queue for r-stage tandem queueing system with Poisson arrivals where the r service times of the same customer at different stations have an arbitrary joint distribution. He showed that for $r = 2$, when service times are positively quadrant dependent, the mean waiting time in queue is greater than when service times are independent. He pointed out that both the results and derivation of some equations in Choo and Conolly [11] are incorrect and simulation results in Mitchell et al. [60] for the same model shed no light on the conflicting results in Choo and Conolly [11] and Pinedo and Wolff [87].

Finally, we review the literature on queueing models in which arrival processes depend upon service processes.

Though admittedly the assumption of independence between arrival and service patterns in queueing models results in a simpler analysis of the processes most commonly studied, it is by far less justifiable from the viewpoint of modelling realistic queueing phenomena. In practice some kind of relationship often exists between arrival and service mechanisms. There are essentially two ways of building models which do not include this basic assumption: either the arrival and/or service parameters are managed to vary with the present number of customers, or consecutive interarrival and service times are not to be statistically independent. The former implies a semi-Markov process for the system state, in the simpler case a generalized Markovian birth and death process for the arrivals and services. The latter means assuming a $2k$-dimensional random variable with positive correlation between components for k consecutive interarrival intervals and the corresponding service times. While numerous special cases of the former exist in the literature, the latter seem to have by and large been ignored.

Conolly [12] and Conolly and Hadidi [13] proposed a new model in which the service time S of any customer is considered to be directly proportional to the interarrival interval T preceding his arrival.

Mitchell and Paulson [61] in a simulation study assumed, for the first time, a kind of stochastic dependence between T and S. Their model is a generalization of the $M/M/1$ in which a correlated pair (T, S) of random variables is now associated with each arriving customer. This pair follows the bivariate exponential distribution which is the so-called Wicksell-Kibble distribution as the same as used in Mitchell et al. [60]. For more details on the concept of correlation and the properties of its bivariate density see [14].

A simulation study of the waiting time process of such a model is reported in Mitchell and Paulson [61] while a mathematical analysis of the customer's waiting time has been given by Conolly and Choo [14] and by Hadidi [31]. Furthermore Hadidi [32] used formulae for the waiting time distribution and its moments, so far obtained, to examine the sensitivity of this distribution to the value of the correlation coefficient.

Langaris [53] obtained a closed-form expression for the Laplace transform of the joint probability and probability density function of the busy period duration and the number of customers served in it.

In those papers the authors considered a single server queue and assumed that the interarrival and service times take the class of Wicksell-Kibble bivariate exponential distribution which is defined using Bessel functions. The Laplace transform of the waiting time distribution is obtained in terms of the solution of some recursive equations, hence it is hard to derive the effect of the correlation.

Chao [9] assumed that the interarrival and service times have the class of bivariate exponential distributions defined by Marshall and Olkin [58]. He showed that the customer waiting time is monotonically decreasing in the dependency in increasing convex ordering sense.

All the above studies considered a single server queue, but Langaris [52] considered the same correlated queue as treated in Langaris [53] with infinitely many servers and derived explicitly formulae concerning the system state probabilities and the output process.

1.2. Review of Queueing Systems with Several Types of Customers

When describing the queueing model of a computer system, a manufacturing system and a telecommunication system, the concept of customer types is very useful. In a computer system, customer types are called as job classes or task classes. Cpu-bound jobs (e.g., numerical analysis, batch, and production jobs), I/O-bound jobs (e.g., interactive, spreadsheet, editing, and game-playing jobs), and mixed jobs (e.g., student, database, and maintenance jobs) represent a few of the common customer classes. In manufacturing system, examples include production facilities which manufacture batch orders for a number of distinct products with the same equipment and/or operators. Often, different service level requirements and/or holding cost rates apply to different items, so that significantly different economic consequences result from the delays or sojourn times experienced by the various items. In a telecommunication system, heterogeneous data types (e.g., interactive message, computer outputs, file transfer, facsimile, etc.) complete with voice for the limited availability for shared transmission equipment, e.g., buses in a local area network or frequency bands in a satellite channel.

Ancker and Gafarian [3] discussed a single server queueing system for m different types of customers having independent Poisson arrivals and exponential service times. Neuts [70] studied an M/G/1 queue with m types of customers in which the server is assumed to expend a random length time in change-over from one type of customer to another.

A multiserver queueing system in which the customers are of several different types have been analyzed in Homma and Fujisawa [36], Kotiah and Slater [49], Slater and Kotiah [96], Smit [97], and Federgruen and Groenevelt [25].

Shioyama [94] studied a two stage tandem queue in which two types of customers are first served at the first stage server and subsequently proceed to the queue at the server corresponding to their types in the second stage.

A closed queueing network system in which the customers are of several different classes have been studied by Kelly [44], Lavenberg and Reiser [57], Bronshtein et al. [7], and Dowdy et al. [20].

Retrial queueing systems are characterized by the feature that arriving calls who find the server busy join the retrial group to try again for their requests in random order and at random intervals. Retrial queues have been widely used to model many problems in telephone switching systems, computer and communication systems. Most retrial queues deal with one type of calls. But there are some practical models which deal with several types of calls. Choi et al. [10] explained two examples as follows.

One example is a telephone switching system (Falin et al. [24]). In modern telephone exchanges, subscriber lines are usually connected to the so-called subscriber line modules. These modules serve both incoming and outgoing calls. An important difference between these types of calls lies in the fact that in the case of blocking due to all channels busy in the module, outgoing calls can be queued, whereas incoming calls get busy signal and must be retried in order to establish the connection. As soon as the channel is free, an outgoing call, if present, occupies the channel immediately. Therefore incoming calls may not establish the connection as long as there are outgoing calls waiting. This fact implies that outgoing calls have non-preemptive priority over incoming calls.

Another example is a mobile cellular radio communication system (Yoon and Un [103]). For an effect use of frequency channels, the service area is divided into a certain number of cells so that the base station in each cell can reuse the channels used in the other cells at the same time. The base station in a cell handles two types of calls. One type is the call initiated in its cell (originating cell). A subscriber with a blocked cell usually reinitiates his attempt after random time. The other type arises when a subscriber holding the line enters the cell from adjacent cells (handoff call). If the base station fails to assign an idle channel until the subscriber gets out of the overlap region of the cells, he suffers from a breakdown during the conversation. The degradation of the quality of the telephone service caused by such a breakdown is more serious than caused by a blocking of an originating call. Thus the base station may give priority to a handoff call by assigning a queue. In the mobile cellular radio communication, the loss of handoff call and the time needed for an originating call to get a channel are the important factors for the quality of service.

A retrial queueing system with two types of customers have been analyzed in Kulkarni [51], and Choi et al. [10].

1.3. Multivariate Exponential Distribution of Marshall and Olkin

There have been several formulations of bivariate exponential distributions. These include the distributions of Gumbel [29], Freund's bivariate extension [26], the distributions of Downton [21] and Hawkes [34]. Marshall and Olkin [58] have proposed a very important bivariate exponential distribution, the BVE. One of the derivations of the BVE is based on a generation of the complete memoryless property of the univariate exponential distribution. This generalization is given in the following definition.

Definition 1.1. A bivariate random variable (X, Y) is said to have the loss of memory property (LMP) iff

$$\overline{F}(s_1 + t, s_2 + t) = \overline{F}(s_1, s_2)\overline{F}(t, t) \quad \text{for } s_1, s_2, t \geq 0, \tag{1.1}$$

where

$$\overline{F}(s, t) = P(X > s, Y > t). \tag{1.2}$$

Marshall and Olkin showed that by assuming exponential marginals and the LMP, the BVE is obtained. In the first place, Marshall and Olkin proposed the BVE as a dependent life time distribution in reliability theory. Another derivation of the BVE is given by the following fatal shock model (see Barlow and Proschan [5]).

Suppose three independent sources of shocks are present in the environment. A shock from source 1 destroys component 1; it occurs at a random time U_1, where $P(U_1 > t) = e^{-\mu_1 t}$. A shock from source 2 destroys component 2; it occurs at a random time U_2, where $P(U_2 > t) = e^{-\mu_2 t}$. Finally, a shock from source 3 destroys both components; it occurs at a random time U_{12}, where $P(U_{12} > t) = e^{-\mu_{12} t}$. Thus the random life length T_1 of component 1 satisfies

$$T_1 = \min(U_1, U_{12}), \tag{1.3}$$

while the random life length T_2 of component 2 satisfies

$$T_2 = \min(U_2, U_{12}), \tag{1.4}$$

Hence the joint survival probability

$$\overline{F}(t_1, t_2) = P(T_1 > t_1, T_2 > t_2) = \exp\{-\mu_1 t_1 - \mu_2 t_2 - \mu_{12} \max(t_1, t_2)\} \tag{1.5}$$

for $t_1 \geq 0, t_2 \geq 0$. The joint distribution $F(t_1, t_2)$ with survival probability given by (1.5) is called the bivariate exponential distribution (BVE). The nonfatal shock model also yields the BVE, as well as fatal.

The BVE has exponential marginal distributions with survival probabilities given by:

$$\overline{F_1}(t_1) = P(T_1 > t_1) = \exp\{-(\mu_1 + \mu_{12}) t_1\} \quad \text{for } t_1 \geq 0,$$

$$\overline{F_2}(t_2) = P(T_2 > t_2) = \exp\{-(\mu_2 + \mu_{12}) t_2\} \quad \text{for } t_2 \geq 0. \tag{1.6}$$

Similarly, we shall treat the multivariate exponential distribution (abbreviated as MVE) of Marshall and Olkin. To fix ideas, consider first an extension of fatal shock model to a three-component system.

Assume the Poisson $Z_i(t)$ with rate μ_i governs the occurrence of shocks fatal to component i for $i = 1, 2, 3$, the Poisson process $Z_{ij}(t)$ with rate μ_{ij} governs the occurrence of shocks fatal to components i and j simultaneously for $1 \le i < j \le 3$, and the Poisson process $Z_{123}(t)$ with rate μ_{123} governs the occurrence of shocks fatal to all three components simultaneously. Assume all the Poisson processes are independent. Let T_i denote the life length of component i for $i = 1, 2, 3$. Then the joint survival probability

$$\begin{aligned}\overline{F}(t_1, t_2, t_3) &= P(T_1 > t_1, T_2 > t_2, T_3 > t_3) \\ &= P[Z_i(t_i) = 0, Z_{ij}\{\max(t_i, t_j)\} = 0, \\ &\quad Z_{123}\{\max(t_1, t_2, t_3)\} = 0], \qquad 1 \le i < j \le 3. \end{aligned} \tag{1.7}$$

Thus

$$\begin{aligned}\overline{F}(t_1, t_2, t_3) =& \exp\{-\mu_1 t_1 - \mu_2 t_2 - \mu_3 t_3 \\ & - \mu_{12} \max(t_1, t_2) - \mu_{13} \max(t_1, t_3) \\ & - \mu_{23} \max(t_2, t_3) - \mu_{123} \max(t_1, t_2, t_3)\}. \end{aligned} \tag{1.8}$$

By similar arguments we obtain the n-dimensional *multivariate exponential distribution* (MVE) with joint survival probability:

$$\begin{aligned}\overline{F}(t_1, \cdots, t_n) =& \exp\{-\sum_{i=1}^{n} \mu_i t_i - \sum_{i<j} \mu_{ij} \max(t_i, t_j) \\ & - \sum_{i<j<k} \mu_{ijk} \max(t_i, t_j, t_k) - \cdots - \mu_{12\cdots n} \max(t_1, \cdots, t_n)\}. \end{aligned} \tag{1.9}$$

By setting $t_i = 0$, we obtain an $(n-1)$-dimensional MVE in the remaining variables. By iterating this process, we see that the marginal distributions of all orders are MVE; in particular, the two-dimensional marginals are BVE, and the one-dimensional marginals are exponential.

In the bivariate case, the BVE results when shocks are nonfatal, as well as fatal. In a similar fashion, the MVE may be obtained when shocks are nonfatal.

On the other hand, Assaf et al. [4] showed that the BVE is a bivariate phase type (abbreviated as BPH) distribution and its multivariate extension is a multivariate phase type (abbreviated as MPH). Raftery [88] introduced a continuous multivariate exponential distribution and O'Cinneide and Raftery [80] showed that it is MPH and derived its MPH representation.

Now, we derive the standard MPH representation of the MVE and get the BVE for the simplest case.

The random vector $\mathbf{S} = (S_1, \cdots, S_n)$ is said to have the multivariate exponential distribution (MVE) of Marshall and Olkin if there exist independent exponential random variables $X_1, \cdots, X_k$ such that for $i = 1, \cdots, n$, $S_i = \min_{j \in J_i} X_j$ where $J_i \subset \{1, \cdots, k\}$ (see Esary and Marshall [23]).

Assaf et al. [4] first formulated a multivariate phase type (MPH) distribution in the following way. Suppose $\{V(t) : t \geq 0\}$ is a regular Markov chain with finite state-space E. Let $\Gamma_1, \cdots, \Gamma_n$ be n non-empty subsets of E such that once V enters Γ_i it never leaves. Suppose that $\bigcap_{i=1}^{n} \Gamma_i$ consists of one state Δ into which absorption is certain. Let β be an initial probability vector on E which puts all its mass on states in $E \setminus \{\Delta\}$.

The infinitesimal generator Q of V is of the form

$$\mathbf{Q} = \begin{bmatrix} \mathbf{T} & -\mathbf{Te} \\ \mathbf{0} & 0 \end{bmatrix}, \tag{1.10}$$

where $\mathbf{T}$ is a square matrix, $\mathbf{e}$ is a column vector of ones, and $\mathbf{0}$ is a column vector of zeros. Define $Y_i = \inf\{t : V(t) \in \Gamma_i\}$ $(i = 1, \cdots, n)$. Then the distribution of $(Y_1, \cdots, Y_n)$ is MPH.

Assaf et al. [4] showed that since each X_i is PH, S_i is PH and, hence, the MVE is MPH. It is of interest to give the standard MPH representation of the MVE, using the results of Assaf et al. [4].

To derive the standard MPH representation of the MVE, we specify explicitly the ingredients $E, \Gamma_1, \cdots, \Gamma_n, \mathbf{T}$ and β. The state-space is $E = \{1, \cdots, m, \Delta\}$, with 2^n elements $\mathbf{q} = (q_1, \cdots, q_n)$, where $q_i \in \{0, 1\}$. We have $\Gamma_i = \{\mathbf{q}, \Delta\}$ $(i = 1, \cdots, n)$, where $q_i = 0$, so that $\{\Delta\} = \bigcap_{i=1}^{n} \Gamma_i$. $\mathbf{T}$ has the following block partitioned structure:

$$\mathbf{T} = \begin{bmatrix} d_0 & \mathbf{B_1} & \mathbf{B_2} & \mathbf{B_3} & \cdots & \mathbf{B_k} & \cdots & \mathbf{B_l} & \cdots & \mathbf{B_{n-1}} \\ & \mathbf{D_1} & \mathbf{C_{12}} & \mathbf{C_{13}} & \cdots & \mathbf{C_{1k}} & \cdots & \mathbf{C_{1l}} & \cdots & \mathbf{C_{1n-1}} \\ & & \mathbf{D_2} & \mathbf{C_{23}} & \cdots & \mathbf{C_{2k}} & \cdots & \mathbf{C_{2l}} & \cdots & \mathbf{C_{2n-1}} \\ & & & \mathbf{D_3} & & & & & & \\ & & & & & \vdots & & \vdots & & \vdots \\ & & & & & \mathbf{D_k} & \cdots & \mathbf{C_{kl}} & \cdots & \mathbf{C_{kn-1}} \\ & & & & & & & \mathbf{D_l} & \cdots & \mathbf{C_{ln-1}} \\ & & & & & & & & & \vdots \\ & & & & & & & & & \mathbf{C_{n-2n-1}} \\ & & & & & & & & & \mathbf{D_{n-1}} \end{bmatrix}, \tag{1.11}$$

where

$$d_0 = -\left(\sum_{i=1}^{n} \mu_i + \sum_{i<j}^{n} \mu_{ij} + \sum_{i<j<k}^{n} \mu_{ijk} + \cdots + \mu_{123\cdots n} \right), \tag{1.12}$$

and all the unmarked entries are zeros.

The submatrices are defined as below. The dimensionality of $\mathbf{B}$ is $1 \times \binom{n}{l}$ $(1 \leq l \leq n-1)$, $\mathbf{C}$ is $\binom{n}{k} \times \binom{n}{l}$ $(1 \leq k \leq n-2, 2 \leq l \leq n-1)$ and $\mathbf{D}$ is $\binom{n}{k} \times \binom{n}{k}$ $(1 \leq k \leq n-1)$. If $\mathbf{q} = (q_1, \cdots, q_n)$ and $\mathbf{r} = (r_1, \cdots, r_n)$ are two states in $E \setminus \{\Delta\}$, we denote by $b_{\mathbf{qr}}$, $c_{\mathbf{qr}}$ and $d_{\mathbf{qr}}$ the corresponding element of submatrices $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$, respectively. $b_{\mathbf{qr}}$, $c_{\mathbf{qr}}$ and $d_{\mathbf{qr}}$ will be zero unless one of the following holds:

(i) $r_{i_1} = r_{i_2} = \cdots = r_{i_u} = 0$ $(1 \leq u \leq l)$.

Then

$$b_{\mathbf{qr}} = \mu_{i_1 i_2 \cdots i_l}, \qquad 1 \leq l \leq n-1. \tag{1.13}$$

(ii) There exists u such that $q_{i_u} = r_{i_u} = 0$ $(1 \leq u \leq k)$, and $q_{j_v} \neq r_{j_v}, r_{j_v} = 0$ $(1 \leq v \leq l-k)$.

Then

$$c_{\mathbf{qr}} = \mu_{j_1 \cdots j_{l-k}} + \sum_{u=1}^{k} \mu_{i_u j_1 \cdots j_{l-k}} + \mu_{i_1 \cdots i_k j_1 \cdots j_{l-k}}, \qquad 1 \leq k \leq n-2, 2 \leq l \leq n-1. \tag{1.14}$$

(iii) There exists u such that $q_{i_u} = r_{i_u} = 0$ $(1 \leq u \leq k)$.

Then

$$d_{\mathbf{qr}} = d_0 + \sum_{u=1}^{k} \mu_{i_u} + \sum_{u<v}^{k} \mu_{i_u i_v} + \cdots + \mu_{i_1 \cdots i_k}, \qquad 1 \leq k \leq n-1, \tag{1.15}$$

where $i_u \subset \{1, \cdots, n\}$ and $j_v \subset \{1, \cdots, n\}$.

To define β, let $\mathbf{Y} = (Y_1, \cdots, Y_n)$ be a random vector taking values in $\{1, \cdots, m\}$ and $p_{j_1 \cdots j_n} = P[Y_1 = j_1, \cdots, Y_n = j_n]$, where j_h ranges over $1, \cdots, m$ for each $h = 1, \cdots, n$. Thus if $\beta_{\mathbf{q}}$ is the element of β corresponding to the state $\mathbf{q}$, we have

$$\beta_{\mathbf{q}} = \begin{cases} p_{q_1 \cdots q_n} & \text{if } q_i \neq 0 \ (i = 1, \cdots, n), \\ 0 & \text{otherwise.} \end{cases} \tag{1.16}$$

This completes the specification of the standard MPH representation of the MVE.

Consider the simplest, bivariate case, where $n = 2$ and $m = 3$. Then the state-space consists of the four elements $(1,1), (1,0), (0,1)$ and $(0,0)$, where $\Gamma_1 = \{(0,1),(0,0)\}$, and $\Gamma_2 = \{(1,0),(0,0)\}$, so that $\Delta = (0,0)$. Then $\mathbf{T}$ is 3×3 matrix

$$\mathbf{T} = \begin{bmatrix} -(\mu_1 + \mu_2 + \mu_{12}) & \mu_2 & \mu_1 \\ 0 & -(\mu_1 + \mu_{12}) & 0 \\ 0 & 0 & -(\mu_2 + \mu_{12}) \end{bmatrix}. \tag{1.17}$$

The initial distribution may be written in the form $\beta = (1,0,0,0)$. From the results of Assaf et al. [1], $\overline{F}(t_1, t_2) = P(Y_1 > t_1, Y_2 > t_2)$ has the following closed form

$$\begin{aligned} \overline{F}(t_1,t_2) &= \alpha\, e^{\mathbf{T}\ t_2}\ \mathrm{g}_2\, e^{\mathbf{T}\ (t_1-t_2)}\ \mathrm{g}_1\ \mathbf{e} \qquad \text{if} \quad t_1 \geq t_2 \geq 0, \\ &= \alpha\, e^{\mathbf{T}\ t_1}\ \mathrm{g}_1\, e^{\mathbf{T}\ (t_2-t_1)}\ \mathrm{g}_2\ \mathbf{e} \qquad \text{if} \quad t_2 \geq t_1 \geq 0, \end{aligned} \tag{1.18}$$

where

$$\alpha = (1,0,0), \quad \mathrm{g}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathrm{g}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \text{and } \mathbf{e} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{1.19}$$

They yield

$$\begin{aligned} \overline{F}(t_1,t_2) &= e^{-(\mu_1+\mu_{12})\ t_1 - \mu_2 t_2} \qquad \text{if} \quad t_1 \geq t_2 \geq 0, \\ &= e^{-\mu_1 t_1 - (\mu_2+\mu_{12})\ t_2} \qquad \text{if} \quad t_2 \geq t_1 \geq 0, \end{aligned} \tag{1.20}$$

consequently we get

$$\overline{F}(t_1,t_2) = e^{-\mu_1 t_1 - \mu_2 t_2 - \mu_{12}\ \max\ (t_1,t_2)}, \tag{1.21}$$

which agrees with the bivariate exponential distribution (BVE) obtained in (1.5).

1.4. Purpose of the Thesis

The purpose of this thesis is to provide explicit solutions for three types of queueing models as follows. First, we pay attention to relaxing the assumption of independent service times out of three sets of independence assumptions in queueing theory and analyze both multiserver and tandem queueing systems with correlated service times. To relax independence of service process, we introduce the multivariate exponential distribution of Marshall and Olkin described in the previous section as the joint service time distribution. We investigate the effect of correlated service times both for several kind of multiserver and tandem queueing models comparing the independent case.

Secondly, we analyze tandem queueing systems with several types of customers. An extension from single to several types of customers is one simple assumption of a tandem queueing model, but we study as a research of a tandem queueing system by considering some kind of matters which are concerned with tandem queueing systems, such as service order, blocking and switching rule etc..

Thirdly, we investigate a multiserver queueing system with additional service channels. As mentioned above, there exists the assumption of independence between arrival and service patterns in queueing theory. One of ways to relax this assumption is to manage to vary service parameters with the present number of customers under the fixed number of channels . We change this point of view and analyze a queueing model whose number of channels are managed to vary with the present number of customers under fixed service parameters.

1.5. Outline of the Thesis

This thesis consists of seven chapters. Chapter 2, 3 and 4 are concerned with queueing systems with correlated service times. Chapter 5 discusses queueing systems with several types of customers. Chapter 6 treats a queueing system with additional service channels. In what follows, we shall summarize the contents of each chapter.

In Chapter 2, we consider a multiserver queueing system with correlated service times whose distribution is the multivariate exponential distribution of Marshall and Olkin discussed in the previous section of Chapter 1. We treat both jockeying and no jockeying cases for this model. The steady state probabilities and the waiting time distribution are derived for the no jockeying case using a generating function approach, whereas the steady state probability vector is obtained for the jockeying case using a matrix-geometric approach.

In Chapter 3, we analyze a tandem queueing system with correlated service times whose distribution is assumed to follow the multivariate exponential distribution of Marshall and Olkin. We investigate both ordinary and commutative tandem queueing systems. In ordinary tandem queueing systems, customers are processed through the ordered sequence of stages, whereas, in commutative tandem queueing systems, customers are served by an empty channel in disregard of the ordered sequence of stages. It has already been shown that the mean queue length of a commutative tandem model is smaller than that of an ordinary tandem model in the case of independent exponential service times. In the case of correlated service times, we study the monotonicity of both the mean number of customers for an ordinary tandem model and the throughput for a commutative tandem model with respect to the correlation coefficient of service times.

Finally, considering the results obtained in Chapter 2 and the former section of Chapter 3, we study the effect of correlated service times on a two-server parallel and two-stage tandem queueing system, respectively.

In Chapter 4, we consider an interchangeable parallel two-stage tandem queueing system in which each stage consists of two channels in parallel. The service times of two channels in the first and second stage are assumed to be the bivariate exponential distribution of Marshall and Olkin. Interchangeable means that a customer who finished the first stage service can enter both of the second stage channels. On the other side, the operation that after completion of the first stage service, each customer can enter only the second stage channel which is located on the same line as his first stage channel is called as ordinary in this chapter. We calculate the throughput for four models in which either of an interchangeable or ordinary operation and either of correlated or independent service times are combined, and compare with these four throughput one another.

In Chapter 5, we analyze a tandem queueing system with several types of customers in which the service distribution is exponential or general. Two cases of finite and infinite intermediate buffers are treated in this chapter. The former case causes the blocking phenomenon which is important in tandem queueing systems. We calculate the mean queue length both for a two-stage ordinary and commutative tandem queueing system with several types of customers, and establish the concavity or convexity of performance measures such as the mean number of busy stations for a two-stage ordinary tandem queueing system with no queues ahead of the first stage. The latter is used for modelling operating systems in computer and switching systems. For this purpose, the assumption that the system is served by a single server is added. We calculate the mean number of customers for a multi-stage tandem queueing system with zero switchover time, which is served according to a exhaustive service.

In Chapter 6, we consider a multiserver queueing system with additional service channels, whose service times is exponentially distributed. We give a explicit expression for the steady state probabilities.

Finally, in Chapter 7, we summarize the results obtained in this thesis and discuss further directions of the analysis for queueing systems with correlated service times.

Chapter 2.

MULTISERVER QUEUEING SYSTEMS WITH CORRELATED SERVICE TIMES

2.1. Introduction

This chapter deals with a correlated multiserver queueing system in which their service distribution is the multivariate exponential distribution of Marshall and Olkin discussed in Chapter 1.

Nishida et al. [74] considered a two-server queueing system in which the interarrival distribution is exponential and their service distribution is the bivariate exponential distribution of Marshall and Olkin with the same service rate and obtained the steady state probabilities.

In Section 2.2, as an extension of the above system, we shall consider a multiserver queueing system which has arbitrary c servers, and show how to get the steady state probabilities, and calculate them explicitly in the case $c = 3$. Furthermore, for this model, we shall derive the queueing time distribution and the waiting time distribution, and show that the consequent result agrees with the so-called Little's formula.

On the other hand, jockeying among queues with servers operating in parallel is an interesting topic in queueing theory. The problem has been studied by Haight [33] and

Koenigsberg [47] for two-server case, and Disney and Mitchell [18], Elsayed and Bastani [22], and Kao and Lin [38] for multiserver case. All these authors considered only models with independent service times. Pivotal to these papers was the assumption of instantaneous jockeying under exponential queue–namely, jockeying when deemed advantageous would occur immediately after service completion. Kao and Lin [38] proposed a matrix-geometric formulation of the jockeying problem considered by Elsayed and Bastani [22]. By exploiting the structure of the model, they developed a simple procedure for computing the stationary probability vector underlying continuous time Markov chain.

In Section 2.3, we shall extend the work of Kao and Lin [38] to compute the stationary probability vector of the correlated multiserver case where the service distribution is the multivariate exponential distribution of Marshall and Olkin. In Section 2.4, we give a summary of Chapter 2.

2.2. Correlated Multiserver Queueing System with Random Input

2.2.1. Model

Fig. 2-1 shows a multiserver queueing system with correlated service times. As in the case of ordinary $M/M/c(N)$, it is assumed that the successive customers arrive according to a Poisson law with parameter $\lambda > 0$, and the system is able to hold a maximum N of customers, and the ordinary queue discipline is being employed. Assume the Poisson process $Z_i(t)$ with rate μ governs the occurrence of service completion to server i for $1 \leq i \leq c$, the Poisson process $Z_{ij}(t)$ with rate α governs the occurrence of service completion to server i and j simultaneously for $1 \le i \le c-1, i+1 \le j \le c$, and more than two simultaneous service completion to servers cannot occur since it is not realistic. The joint distribution of service times X_i $(1 \le i \le c)$ is given from (1.9) as follows:

$$Pr[X_1 > x_1, X_2 > x_2, \cdots, X_c > x_c] = \exp[-\mu \sum_{i=1}^{c} x_i - \alpha \sum_{j=i+1}^{c} \sum_{i=1}^{c-1} \max(x_i, x_j)], \quad (2.1)$$

which is the service distribution we shall adopt throughout this section. The correlation coefficient of X_m and X_n, $m \neq n$, is

$$\rho[X_m, X_n] = \alpha/[2\mu + (2c-3)\alpha]. \quad (2.2)$$

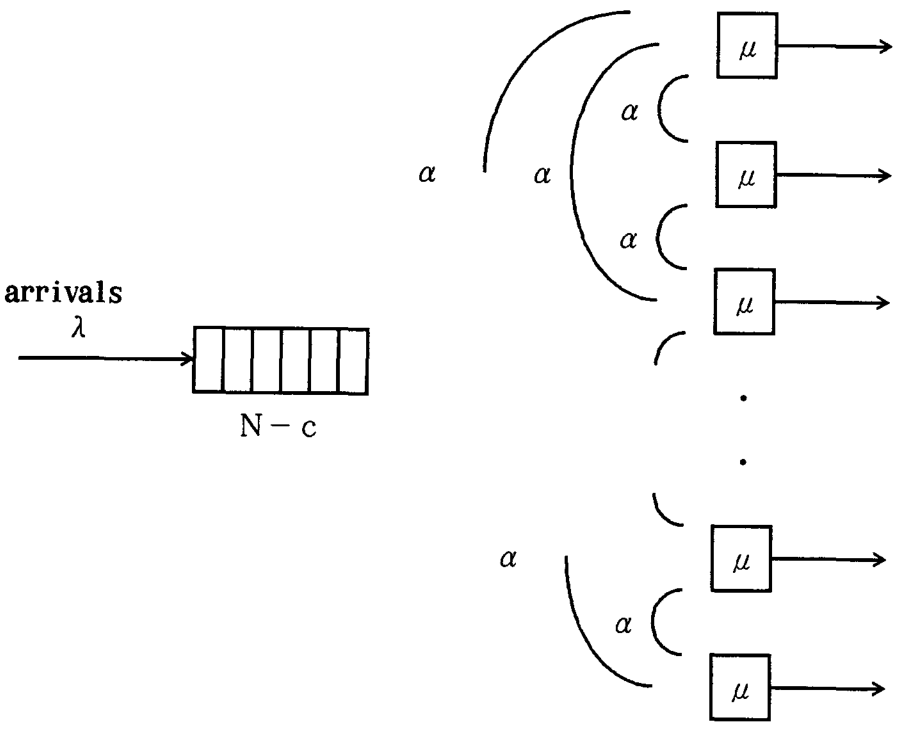

Fig. 2-1. The M/MVE/c(N) queueing system.

Calculating the r-dimensional marginals from (2.1), we obtain

$$\exp[-\{\mu + (c-r)\alpha\} \sum_{i=1}^{r} x_{k_i} - \alpha \sum_{j=i+1}^{r} \sum_{i=1}^{r-1} \max(x_{k_i}, x_{k_j})]. \tag{2.3}$$

Especially the one-dimensional marginals of the MVE are exponential. On the other hand, the r-dimensional random variable $\min(X_1, X_2, \cdots, X_r)$ has the following distribution from (2.3):

$$Pr[\min(X_1, X_2, \cdots, X_r) > x] = \exp[-\frac{1}{2} r\{2\mu + (2c - r - 1)\alpha\}x]. \tag{2.4}$$

The fact that the r-dimensional random variable $\min(X_1, X_2, \cdots, X_r)$ is exponential is an important property, which will be useful in later discussion.

2.2.2. Steady State Probabilities

Let $P_n(t)$ denote the probability that the number of customers in the system is n at the time $t \geq 0$. Then we may write down the differential-difference equations for this system as follows:

$$\begin{aligned}
P_0'(t) &= -\lambda P_0(t) + a_1 P_1(t) + b_2 P_2(t), \\
P_n'(t) &= \lambda P_{n-1}(t) - (\lambda + a_n) P_n(t) + (a_{n+1} - b_{n+1}) P_{n+1}(t) \\
&\quad + b_{n+2} P_{n+2}(t), \qquad 1 \leq n \leq c-2, \\
P_{c-1}'(t) &= \lambda P_{c-2}(t) - (\lambda + a_{c-1}) P_{c-1}(t) + (a_c - b_c) P_c(t) + b_c P_{c+1}(t), \\
P_n'(t) &= \lambda P_{n-1}(t) - (\lambda + a_c) P_n(t) + (a_c - b_c) P_{n+1}(t) + b_c P_{n+2}(t), \quad c \leq n \leq N-2, \\
P_{N-1}'(t) &= \lambda P_{N-2}(t) - (\lambda + a_c) P_{N-1}(t) + (a_c - b_c) P_N(t), \\
P_N'(t) &= \lambda P_{N-1}(t) - a_c P_N(t),
\end{aligned} \tag{2.5}$$

where

$$a_n = \frac{1}{2}n[2\mu + (2c - n - 1)\alpha], \quad b_n = \frac{1}{2}n(n-1)\alpha, \quad 1 \le n \le c. \tag{2.6}$$

From (2.5), one can readily obtain the following set of difference equations for the steady state probabilities $p_n = \lim_{t\to\infty} P_n(t)$.

$$\begin{aligned} &-\lambda p_n + a_{n+1}p_{n+1} + b_{n+2}p_{n+2} = 0, \quad 0 \le n \le c-2, \\ &-\lambda p_n + a_c p_{n+1} + b_c p_{n+2} = 0, \quad c-1 \le n \le N-2, \\ &-\lambda p_{N-1} + a_c p_N = 0. \end{aligned} \tag{2.7}$$

Also the normalization condition yields

$$\sum_{n=0}^{N} p_n = 1. \tag{2.8}$$

In order to solve (2.7) under the condition (2.8), we shall define the generating function

$$F(z) = \sum_{n=0}^{N} p_n z^n. \tag{2.9}$$

Multiplying the equations in (2.7) successively by $1, z, \cdots, z^N$ and adding them up yield

$$F(z) = [\lambda p_N z^{N+2} - (\mu z + b_c)z^{c-1}p_{c-1} - (a_c z + b_c)z^{c-2}p_{c-2} + f(z)\sum_{n=0}^{c-3} z^n p_n]/[f(z)], \tag{2.10}$$

where

$$f(z) = \lambda z^2 - a_c z - b_c. \tag{2.11}$$

The condition (2.8) gives

$$F(1) = 1,$$

so that we get from (2.10),

$$-\lambda p_N + (\mu + b_c)p_{c-1} + (a_c + b_c)p_{c-2} + (a_c + b_c - \lambda)\sum_{n=0}^{c-3} p_n = a_c + b_c - \lambda. \tag{2.12}$$

The denominator on the right-hand side of (2.10) has two real zeros, namely,

$$\xi = [a_c - (a_c^2 + 4\lambda b_c)^{\frac{1}{2}}]/(2\lambda), \quad \zeta = [a_c + (a_c^2 + 4\lambda b_c)^{\frac{1}{2}}]/(2\lambda). \tag{2.13}$$

Since the generating function $F(z)$ is regular on z-plane, the zeros of the numerator and the denominator of the right-hand side of (2.10) must coincide with each other. Therefore, it follows that

$$\lambda p_N \xi^{N-c+3} - (\mu\xi + b_c)p_{c-1} - \lambda\xi p_{c-2} = 0, \tag{2.14}$$

and

$$\lambda p_N \zeta^{N-c+3} - (\mu\zeta + b_c)p_{c-1} - \lambda\zeta p_{c-2} = 0, \tag{2.15}$$

where ξ and ζ are given by (2.13).

We notice the numerator of the right-hand side of (2.10) involves the $c+1$ unknowns p_n. In order to determine p_n $(0 \leq n \leq c-1)$ and p_N and thus be able to determine p_n $(c \leq n \leq N-1)$, we require $c+1$ equations involving the $c+1$ unknowns p_n. Now (2.12), (2.14) and (2.15) are three such equations and the first $c-2$ equations in (2.7) are the remaining. These $c-2$ equations are

$$-\lambda p_n + a_{n+1}p_{n+1} + b_{n+2}p_{n+2} = 0, \quad 0 \leq n \leq c-3, \tag{2.16}$$

which is valid only when $c \geq 3$. We treat the case $c = 2$ separately. The case $c = 2$ requires (2.12), (2.14) and (2.15) alone, from which p_0, p_1 and p_N are determined.

Returning to (2.10) we note that the remaining p_n $(c \leq n \leq N-1)$ may be obtained as the coefficients of z^n. To do this, we write down the numerator as follows:

$$[z^2 - (\xi + \zeta)z + \xi\zeta][A_N z^N + A_{N-1}z^{N-1} + \cdots + A_0]. \tag{2.17}$$

Comparing the coefficients of $z^{c+1}, z^{c+2}, \cdots, z^N$, we obtain the difference equations:

$$\xi\zeta A_n - (\xi+\zeta)A_{n-1} + A_{n-2} = 0, \quad c+1 \le n \le N,$$
$$-(\xi+\zeta)A_N + A_{N-1} = 0,$$
$$A_N = \lambda p_N. \tag{2.18}$$

Solving (2.18) and using the relation $A_n/\lambda = p_n$,

$$p_n = p_N(\xi^{N+1-n} - \zeta^{N+1-n})/(\xi-\zeta), \quad c \le n \le N-1, \tag{2.19}$$

is derived, where p_N is obtained as the solution of the previous equations.

We shall investigate two extreme systems in which the capacity of waiting room is zero and infinity, respectively. We set $N = c$ in (2.12), (2.14) and (2.15) and solve these three equations together with (2.16), which results in the case of $M/MVE/c(c)$ queueing system. We next consider the case $N = \infty$. It is first to be noted that

$$|\xi| < \zeta, \tag{2.20}$$

and

$$\zeta > 1 \quad \text{if and only if} \quad \lambda < c[\mu + (c-1)\alpha]. \tag{2.21}$$

Therefore, as $N \to \infty$ in the probabilities $p_n (0 \le n \le N)$ of $M/MVE/c(N)$, supposing $\lambda < c[\mu + (c-1)\alpha]$, we have the steady state probabilities of $M/MVE/c(\infty)$. If, on the other hand, $\lambda \ge c[\mu + (c-1)\alpha]$, then

$$\lim_{N\to\infty} p_n = 0, \quad n = 0, 1, \cdots, \tag{2.22}$$

which implies that the steady state conditions are not satisfied. Also, since the term with p_N in the numerator of the right-hand side in (2.10) vanishes, the generating function of $M/MVE/c(\infty)$ is

$$G(z) = [-(\mu z + b_c)z^{c-1}p_{c-1} - (a_c z + b_c)z^{c-2}p_{c-2}]/[f(z)] + \sum_{n=0}^{c-3} z^n p_n, \tag{2.23}$$

where $f(z)$ is given by (2.11). The above generating function will be used for calculating some average characteristics for $M/MVE/c(\infty)$.

Let us treat the case $c = 3$. From (2.6) and (2.10), the generating function

$$F(z) = \frac{\lambda p_N z^{N+2} - \mu p_2 z^3 - (2\alpha p_2 + 2\mu p_1 + \alpha p_1)z^2 - (3\alpha p_1 + 3\mu p_0 + 3\alpha p_0)z - 3\alpha p_0}{\lambda(z-\xi)(z-\zeta)}, \tag{2.24}$$

where

$$\xi = [3(\mu+\alpha) - \{9(\mu+\alpha) + 12\alpha\lambda\}^{\frac{1}{2}}]/(2\lambda), \tag{2.25}$$

and

$$\zeta = [3(\mu+\alpha) + \{9(\mu+\alpha) + 12\alpha\lambda\}^{\frac{1}{2}}]/(2\lambda). \tag{2.26}$$

Solving (2.12), (2.14), (2.15) and (2.16) for $c = 3$ and substituting the obtained p_N into (2.19), we get

$$\begin{aligned} p_0 &= (6\alpha + 3\mu - \lambda)B_1/B_2, \\ p_1 &= \lambda[3\alpha(\xi^N - \zeta^N) + \mu\xi\zeta(\xi^{N-1} - \zeta^{N-1})]p_0/B_1, \\ p_n &= 3\alpha\lambda(\xi^{N+1-n} - \zeta^{N+1-n})p_0/B_1, \quad 2 \le n \le N, \end{aligned} \tag{2.27}$$

where

$$B_1 = 3\alpha(2\alpha+\mu)(\xi^N - \zeta^N) + \xi\zeta(\xi^{N-1} - \zeta^{N-1})(\mu^2 + 2\alpha\mu - \alpha\lambda), \tag{2.28}$$

and

$$B_2 = (2\alpha + \mu)[3\alpha(6\alpha + 3\mu + 2\lambda)(\xi^N - \zeta^N)$$
$$+ \xi\zeta(\xi^{N-1} - \zeta^{N-1})(6\alpha\mu + 3\mu^2 + 2\mu\lambda - 3\alpha\lambda - \lambda^2)] - 3\alpha\lambda^2(\xi - \zeta). \tag{2.29}$$

Upon setting $N = 3$ in (2.27), (2.28) and (2.29), we obtain

$$p_0 = C_1/C_2,$$
$$p_1 = 3\lambda(3\alpha^2 + 2\mu^2 + 5\alpha\mu + \alpha\lambda)p_0/C_1,$$
$$p_2 = 3\lambda^2(\alpha + \mu)p_0/C_1,$$
$$p_3 = \lambda^3 p_0/C_1, \tag{2.30}$$

where

$$C_1 = 3(2\alpha + \mu)(3\alpha^2 + 2\mu^2 + 5\alpha\mu + \alpha\lambda) + 3\alpha\lambda(\alpha + \mu),$$
$$C_2 = 3(2\alpha + \mu + \lambda)(3\alpha^2 + 2\mu^2 + 5\alpha\mu + \alpha\lambda) + 3\lambda(\alpha + \mu)(\alpha + \lambda) + \lambda^3, \tag{2.31}$$

which results in the case of $M/MVE/3(3)$. Noting (2.20), we find from (2.28) and (2.29) that as $N \to \infty$,

$$\frac{B_1}{B_2} \to \frac{(\mu^2 + 2\alpha\mu - \alpha\lambda)\xi + 3\alpha(2\alpha + \mu)}{(2\alpha + \mu)[(6\alpha\mu + 3\mu^2 + 2\mu\lambda - 3\alpha\lambda - \lambda^2)\xi + 3\alpha(6\alpha + 3\mu + 2\lambda)]} \equiv \frac{D_1}{D_2}. \tag{2.32}$$

Therefore, supposing $\lambda < 3(\mu + 2\alpha)$, we have

$$\tilde{p_0} = \lim_{N\to\infty} p_0 = (6\alpha + 3\mu - \lambda)D_1/D_2,$$
$$\tilde{p_1} = \lim_{N\to\infty} p_1 = \lambda(3\alpha + \mu\xi)\tilde{p_0}/D_1,$$
$$\tilde{p_n} = \lim_{N\to\infty} p_n = 3\alpha\lambda\zeta^{1-n}\tilde{p_0}/D_1, \qquad n \geq 2, \tag{2.33}$$

which are the steady state probabilities of $M/MVE/3(\infty)$. Also, from (2.23), the generating function $G(z)$ of $M/MVE/3(\infty)$ is given by

$$G(z) = \frac{-\mu\tilde{p}_2 z^3 - (2\alpha\tilde{p}_2 + 2\mu\tilde{p}_1 + \alpha\tilde{p}_1)z^2 - (3\alpha\tilde{p}_1 + 3\mu\tilde{p}_0 + 3\alpha\tilde{p}_0)z - 3\alpha\tilde{p}_0}{\lambda(z-\xi)(z-\zeta)}. \tag{2.34}$$

Using the generating function (2.34), the mean number of customers is

$$\begin{aligned} L &= G'(1) \\ &= \frac{(6\mu^2 + 18\alpha^2 + 21\alpha\mu - \mu\lambda)\tilde{p}_2 + 3(\alpha\lambda + 2\mu^2 + 7\alpha\mu + 6\alpha^2)\tilde{p}_1 + 3\lambda(\mu + 3\alpha)\tilde{p}_0}{(\lambda - 3\mu - 6\alpha)^2}. \end{aligned} \tag{2.35}$$

If $\alpha = 0$, the service times of three servers are mutually independent, hence by letting $\alpha \to 0$ in (2.27), (2.28) and (2.29), one will get the results of ordinary $M/M/3(N)$ with independent servers which we shall see below.

We shall write $\xi(\alpha), \zeta(\alpha), B_1(\alpha)$ and $B_2(\alpha)$ respectively for ξ, ζ, B_1 and B_2. From (2.25) and (2.26), it is easily seen that

$$\xi(0) = 0, \quad \xi'(0) = -1/\mu, \quad \zeta(0) = 1/\rho, \tag{2.36}$$

where we put $\rho = \lambda/3\mu$. Hence $B_1(0) = B_2(0) = 0$, but

$$B_2'(0) = -2\mu/\rho^N, \tag{2.37}$$

and

$$B_2'(0) = -3\mu^2(2 + 4\rho + 3\rho^2 - 9\rho^{N+1})/\rho^N, \tag{2.38}$$

from which, employing L'Hospital's rule, we have

$$\hat{p_0} = \lim_{\alpha \to 0} p_0 = \frac{2(1-\rho)}{2 + 4\rho + 3\rho^2 - 9\rho^{N+1}}. \tag{2.39}$$

Similarly,

$$\lim_{\alpha \to 0} p_1 = 3\rho \hat{p}_0, \tag{2.40}$$

$$\lim_{\alpha \to 0} p_n = \frac{9}{2}\rho^n \hat{p}_0, \qquad 2 \le n \le N, \tag{2.41}$$

which agrees with the solution of $M/M/3(N)$ queueing system[28].

2.2.3. Waiting Time Distribution

First, we shall obtain the queueing time distribution of the $M/MVE/c(\infty)$ system. Since the generating function $G(z)$ given by (2.23) is regular in the domain $|z| \le 1$, the root of the numerator and the denominator of the first term of the right-hand side of (2.23) must coincide with each other.

Therefore,

$$(\mu\xi + b_c)\xi^{c-1} p_{c-1} + (a_c\xi + b_c)\xi^{c-2} p_{c-2} = 0. \tag{2.42}$$

From (2.20), (2.21) and (2.42), the generating function (2.23) is

$$G(z) = \frac{\mu(a_c\xi + b_c)z + b_c(\mu\xi + b_c)}{\lambda\zeta(a_c\xi + b_c)} \cdot p_{c-1} \sum_{n=0}^{\infty} (1/\zeta)^n z^{n+c-2} + \sum_{n=0}^{c-3} z^n p_n. \tag{2.43}$$

Comparing the coefficients of $z^n (n \ge c)$ and using the equations (2.13), we obtain the steady state probabilities $p_n (n \ge c)$ as follows.

$$p_n = p_{c-1} \, (1/\zeta)^{n+1-c}, \quad n \ge c, \tag{2.44}$$

where how to get p_{c-1} has been given in Section 2.2.2.

If $\alpha = 0$, the service times are mutually independent, hence by letting $\alpha \to 0$ in (2.13),

$$\lim_{\alpha \to 0} \zeta = \frac{c\mu}{\lambda}, \tag{2.45}$$

so (2.44) yields

$$p_n = p_{c-1} \left(\frac{\lambda}{c\mu} \right)^{n+1-c}, \quad n \geq c, \tag{2.46}$$

which agrees with the solution of ordinary $M/M/c(\infty)$ queueing system [28].

We define the events B_i and B_{ij} in order to derive the queueing time distribution as follows:

B_i $(1 \leq i \leq c)$: the event that only one the i-th channel finishes service,

B_{ij} $(1 \leq i \leq c-1, i+1 \leq j \leq c)$: the event that two the i-th and the j-th channels finish service at the same time.

By the definition of the multivariate exponential distribution (MVE), the events B_i and B_{ij} are mutually independent, and

$$\begin{aligned} &Pr\{\text{the event } B_i \text{ does not occur in } \leq t\} = e^{-\mu t}, \quad 1 \leq i \leq c, \\ &Pr\{\text{the event } B_{ij} \text{ does not occur in } \leq t\} = e^{-\alpha t}, \quad 1 \leq i \leq c-1, i+1 \leq j \leq c. \end{aligned} \tag{2.47}$$

Therefore,

$$\begin{aligned} Pr\{\text{none of channels completes in } \leq t\} &= Pr(\bigcap_{i=1}^{c} \bar{B}_i \bigcap_{j=i+1}^{c} \bigcap_{i=1}^{c-1} \bar{B}_{ij}) \\ &= e^{-({}_cC_1\mu + {}_cC_2\alpha)t}, \end{aligned} \tag{2.48}$$

where $\bar{B}_i$ denotes that the event B_i does not occur in $\leq t$.

Thus, we get

$$Pr\{\text{just one channel completes in} \le t\} = Pr(B_1 \bigcap_{i=2}^{c} \bar{B}_i \bigcap_{j=i+1}^{c} \bigcap_{i=1}^{c-1} \bar{B}_{ij})$$
$$= {}_cC_1(\mu t e^{-\mu t}) e^{-(c-1)\mu t} e^{-{}_cC_2 \alpha t}$$
$$= {}_cC_1 \mu t e^{-({}_cC_1\mu + {}_cC_2\alpha)t}, \quad (2.49)$$

and

$$Pr\{\text{two channels complete in} \le t\} = Pr(B_1^2 \bigcap_{i=2}^{c} \bar{B}_i \bigcap_{j=i+1}^{c} \bigcap_{i=1}^{c-1} \bar{B}_{ij})$$
$$+ Pr\{(B_1 \cap B_2) \bigcap_{i=3}^{c} \bar{B}_i \bigcap_{j=i+1}^{c} \bigcap_{i=1}^{c-1} \bar{B}_{ij}\}$$
$$+ Pr(B_{12} \bigcap_{i=1}^{c} \bar{B}_i \bigcap_{j=i+1, j\neq 2}^{c} \bigcap_{i=1}^{c-1} \bar{B}_{ij}\}$$
$$= [(\frac{1}{2}{}_cC_1 + {}_cC_2)(\mu t)^2 + {}_cC_2(\alpha t)] e^{-({}_cC_1\mu + {}_cC_2\alpha)t}. \quad (2.50)$$

By the same manner,

$$Pr\{n \text{ channels complete in} \le t\}$$
$$= \sum_{i=0}^{[n/2]} \frac{({}_cC_1 \mu t)^{n-2i}}{(n-2i)!} \cdot \frac{({}_cC_2 \alpha t)^i}{i!} \cdot e^{-({}_cC_1\mu + {}_cC_2\alpha)t}. \quad (2.51)$$

Now, let T_q represent the random variable "time spent waiting in the queue" and $W_q(t)$ denote its CDF.

Therefore, from (2.44) we have

$$W_q(0) = Pr\{T_q = 0\}$$
$$= Pr\{c-1 \text{ or less in the system}\}$$
$$= \sum_{n=0}^{c-1} p_n$$

$$= 1 - \sum_{n=c}^{\infty} p_n$$

$$= 1 - \sum_{n=c}^{\infty} p_{c-1}(1/\zeta)^{n+1-c}$$

$$= 1 - \frac{p_{c-1}}{\zeta - 1}. \qquad (2.52)$$

For $T_q > 0$, then, from (2.13), (2.44) and (2.51)

$$W_q(t) = Pr\{T_q \leq t\}$$

$$= 1 - \sum_{m=0}^{\infty} p_{m+c}\, Pr\{m \text{ completions in } \leq t \mid \text{arrival found } m+c \text{ in the system}\}$$

$$= 1 - \sum_{m=0}^{\infty} p_{m+c} \sum_{n=0}^{m} \sum_{i=0}^{[n/2]} \frac{({}_cC_1\mu t)^{n-2i}}{(n-2i)!} \cdot \frac{({}_cC_2\alpha t)^i}{i!} \cdot e^{-({}_cC_1\mu + {}_cC_2\alpha)t}$$

$$= 1 - \frac{p_{c-1}}{\zeta - 1} e^{-\lambda(\zeta-1)t}. \qquad (2.53)$$

If $\alpha = 0$, from (2.45) and (2.46), (2.52) and (2.53) yield

$$\lim_{\alpha \to 0} W_q(0) = 1 - \frac{\lambda p_{c-1}}{c\mu - \lambda}, \qquad (2.54)$$

and

$$\lim_{\alpha \to 0} W_q(t) = 1 - \frac{\lambda p_{c-1}}{c\mu - \lambda} e^{-(c\mu-\lambda)t}, \qquad (2.55)$$

which coincide with the ordinary case [28].

From (2.53), we have the mean queueing time

$$W_q = E[T_q] = \int_0^{\infty} t dW_q(t) = \frac{p_{c-1}}{\lambda(\zeta - 1)^2}. \qquad (2.56)$$

On the other hand, from (2.44) the mean queue length is

$$L_q = \sum_{n=c+1}^{\infty} (n-c)p_n = \frac{p_{c-1}}{(\zeta - 1)^2}. \tag{2.57}$$

Consequently,

$$L_q = \lambda W_q, \tag{2.58}$$

which implies that Little's formula is valid.

Next, we shall direct our attention to the waiting time distribution. The probability that a customer who arrives at $t = 0$ joins in any channel between time y and $y + dy$ is

$$w_q(t)dy, \tag{2.59}$$

where $w_q(t)$ is the queueing time density function. And the probability that a customer who arrives at time y completes a service between time x and $x + dx$ is

$$\{\mu + (c-1)\alpha\}e^{-\{\mu+(c-1)\alpha\}(x-y)}dx, \tag{2.60}$$

since the r-dimensional marginals of the service distribution is given by (2.3). Thus, the probability that a customer who arrives at $t = 0$ departs from the system between time x and $x + dx$ is

$$[\int_0^x \{\mu + (c-1)\alpha\}e^{-\{\mu+(c-1)\alpha\}(x-y)}w_q(y)dy]dx. \tag{2.61}$$

On the other hand, the probability that there is no queue and any arriving customer can immediately receive a service is

$$\sum_{n=0}^{c-1} p_n, \tag{2.62}$$

and at the time the probability that the waiting time is greater than t is

$$e^{-\{\mu+(c-1)\alpha\}t}. \tag{2.63}$$

Consequently, the waiting time distribution $W(t)$ is given as follows.

$$1-W(t)=e^{-\{\mu+(c-1)\alpha\}t}\sum_{n=0}^{c-1}p_n+\int_t^{\infty}[\int_0^x\{\mu+(c-1)\alpha\}e^{-\{\mu+(c-1)\alpha\}(x-y)}w_q(y)dy]dx$$

$$=\left[1+\frac{\{\mu+(c-1)\alpha\}p_{c-1}}{(\zeta-1)\{\lambda(\zeta-1)-\mu-(c-1)\alpha\}}\right]e^{-\{\mu+(c-1)\alpha\}t}$$

$$-\frac{\{\mu+(c-1)\alpha\}p_{c-1}}{(\zeta-1)\{\lambda(\zeta-1)-\mu-(c-1)\alpha\}}e^{-\lambda(\zeta-1)t}. \quad (2.64)$$

If $\alpha=0$, from (2.45) and (2.46), (2.64) yields

$$\lim_{\alpha\to 0}\{1-W(t)\}=\left[1+\frac{\lambda\mu p_{c-1}}{(c\mu-\lambda)\{(c-1)\mu-\lambda\}}\right]e^{-\mu t}-\frac{\lambda\mu p_{c-1}}{(c\mu-\lambda)\{(c-1)\mu-\lambda\}}e^{-(c\mu-\lambda)t}, \quad (2.65)$$

which agrees with that of the ordinary case. From (2.64), we obtain the mean waiting time

$$W=\int_0^{\infty}tdW(t)$$

$$=\frac{1}{\mu+(c-1)\alpha}+\frac{p_{c-1}}{\lambda(\zeta-1)^2}$$

$$=\frac{1}{\mu+(c-1)\alpha}+W_q. \quad (2.66)$$

Now, in order to calculate the mean number of customers L, we shall derive a few equations.

Thus, defining the generating function

$$Q(z)=\sum_{n=0}^{c-3}z^n p_n, \quad (2.67)$$

then the equation given by (2.16):

$$b_{n+2}p_{n+2}+a_{n+1}p_{n+1}-\lambda p_n=0, \quad 0\leq n\leq c-3 \quad (2.68)$$

yields

$$\frac{1}{2}\alpha(1-z)Q''(z)+a_1Q'(z)-\lambda Q(z)$$
$$=-(b_{c-1}p_{c-1}+a_{c-2}p_{c-2})z^{c-3}-b_{c-2}p_{c-2}z^{c-4}. \tag{2.69}$$

Puttung z=1 in (2.69), we have the equation

$$Q'(1)=\{\lambda Q(1)-b_{c-1}p_{c-1}-(a_{c-2}+b_{c-2})p_{c-2}\}/a_1. \tag{2.70}$$

The normalization condition from (2.23) gives the equation

$$Q(1)=1-\frac{(\mu+b_c)p_{c-1}+(a_c+b_c)p_{c-2}}{a_c+b_c-\lambda}. \tag{2.71}$$

Using the equations (2.70) and (2.71), the mean number of customers is

$$\begin{aligned} L &= \sum_{n=0}^{\infty} np_n \\ &= \sum_{n=c+1}^{\infty}(n-c)p_n+\sum_{n=0}^{c} np_n + c\sum_{n=c+1}^{\infty} p_n \\ &= L_q+Q'(1)-cQ(1)-2p_{c-2}-p_{c-1}+c \\ &= L_q+\frac{\lambda}{\mu+(c-1)\alpha}, \end{aligned} \tag{2.72}$$

which implies that Little's formula is valid.

2.3. Correlated Multiserver Queueing System with Instantaneous Jockeying

2.3.1. Model and QBD Formulation

Fig. 2-2 shows a queueing system with c servers and Poisson arrivals of rate $\lambda > 0$. The joint distribution of service times $X_i (1 \le i \le c)$ is assumed to follow the multivariate exponential distribution of Marshall and Olkin.

Each server has a separate waiting line. Customers do not renege or balk. The jockeying rules are:

(i) A new arrival joins any of the empty queues with equal probability.

(ii) If there is no empty queue, the arrival joins the shortest queue.

(iii) If the shortest queue is not unique, the arrival joins any one of them with equal probability.

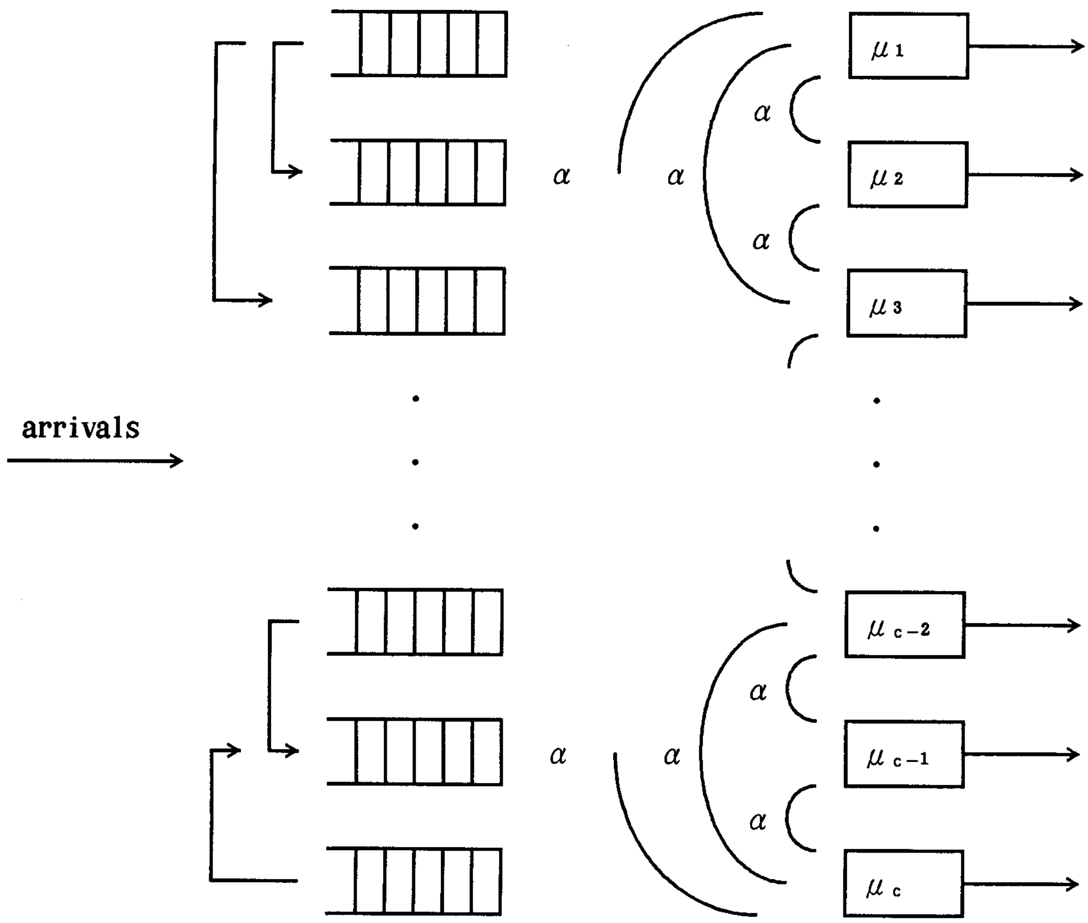

Fig. 2-2. The M/MVE/c(∞) queueing system with instantaneous jockeying.

(iv) A customer in a given queue will jockey to a shorter queue as soon as the difference in the number of waiting customers between the two queues becomes two.

(v) If there are two or more queues with equal length which are shorter than a given queue, then the last unit in the longer queue jockeys to any one of the shorter queues with equal probability.

As Kao and Lin [38], for the notational convenience we consider the case for $c = 3$ in detail. Each state of the QBD process is triplet with each element denoting the number of customers including the customer receiving service in the respective queue.

Let $\mathbf{Q}$ denote the infinitesimal generator of the underlying QBD process. Corresponding to this $\mathbf{Q}$ matrix, we order the states as follows:

$$
\begin{aligned}
&\{(0,0,0)\},\ \{(1,0,0),(0,1,0),(0,0,1),(1,1,0),(1,0,1),(0,1,1),(1,1,1)\},\\
&\{(2,1,1),(1,2,1),(1,1,2),(2,2,1),(2,1,2),(1,2,2),(2,2,2)\},\\
&\{(3,2,2),(2,3,2),(2,2,3),(3,3,2),(3,2,3),(2,3,3),(3,3,3)\},\\
&\{(4,3,3),(3,4,3),(3,3,4),(4,4,3),(4,3,4),(3,4,4),(4,4,4)\},\cdots
\end{aligned}
\tag{2.73}
$$

In the above listing of states, we use pairs of braces to demarcate the submatrices constituting the infinitesimal generator $\mathbf{Q}$ shown below:

$$
\mathbf{Q} = \begin{bmatrix}
\mathbf{A_{01}} & \mathbf{A_{00}} & & & & \\
\mathbf{A_{12}} & \mathbf{A_{11}} & \mathbf{A_0} & & & \\
 & \mathbf{A_2} & \mathbf{A_1} & \mathbf{A_0} & & \\
 & & \mathbf{A_2} & \mathbf{A_1} & \mathbf{A_0} & \\
 & & & \mathbf{A_2} & \mathbf{A_1} & \\
 & & & & \vdots & \ddots
\end{bmatrix},
\tag{2.74}
$$

where $\mathbf{A_{01}}$ is 1×1, $\mathbf{A_{00}}$ is 1×7, $\mathbf{A_{12}}$ is 7×1 and $\mathbf{A_{11}}, \mathbf{A_0}, \mathbf{A_1}, \mathbf{A_2}$ are all 7×7. Defining row vector $\omega = \{\lambda/3, \lambda/3, \lambda/3, 0, 0, 0, 0\}$, we have $\mathbf{A_{01}} = \{-\lambda\}, \mathbf{A_{00}} = \omega$. The matrix

$\mathbf{A_0}$ is a zero matrix except the last row is given by ω. Other matrices are

$$\mathbf{A_{12}} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \alpha \\ \alpha \\ \alpha \\ 0 \end{bmatrix},$$

$$\mathbf{A_{11}} = \begin{bmatrix} -v_1 & 0 & 0 & \lambda/2 & \lambda/2 & 0 & 0 \\ 0 & -v_2 & 0 & \lambda/2 & 0 & \lambda/2 & 0 \\ 0 & 0 & -v_3 & 0 & \lambda/2 & \lambda/2 & 0 \\ u_2-\alpha & u_1-\alpha & 0 & -v_{12} & 0 & 0 & \lambda \\ u_3-\alpha & 0 & u_1-\alpha & 0 & -v_{31} & 0 & \lambda \\ 0 & u_3-\alpha & u_2-\alpha & 0 & 0 & -v_{23} & \lambda \\ \alpha & \alpha & \alpha & \mu_3 & \mu_2 & \mu_1 & -v_{123} \end{bmatrix},$$

$$\mathbf{A_1} = \begin{bmatrix} -v_{123} & 0 & 0 & \lambda/2 & \lambda/2 & 0 & 0 \\ 0 & -v_{123} & 0 & \lambda/2 & 0 & \lambda/2 & 0 \\ 0 & 0 & -v_{123} & 0 & \lambda/2 & \lambda/2 & 0 \\ w_{23} & w_{13} & 0 & -v_{123} & 0 & 0 & \lambda \\ w_{32} & 0 & w_{12} & 0 & -w_{123} & 0 & \lambda \\ 0 & w_{31} & w_{21} & 0 & 0 & -v_{123} & \lambda \\ \alpha & \alpha & \alpha & \mu_3 & \mu_2 & \mu_1 & -v_{123} \end{bmatrix},$$

$$\mathbf{A_2} = \begin{bmatrix} 0 & 0 & 0 & 3\lambda/2 & 3\lambda/2 & 0 & \mu' \\ 0 & 0 & 0 & 3\lambda/2 & 0 & 3\lambda/2 & \mu' \\ 0 & 0 & 0 & 0 & 3\lambda/2 & 3\lambda/2 & \mu' \\ 0 & 0 & 0 & 0 & 0 & 0 & 3\alpha \\ 0 & 0 & 0 & 0 & 0 & 0 & 3\alpha \\ 0 & 0 & 0 & 0 & 0 & 0 & 3\alpha \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{2.75}$$

where

$$u_1 = \mu_1 + 2\alpha,$$

$$u_2 = \mu_2 + 2\alpha,$$

$$u_3 = \mu_3 + 2\alpha,$$

$$v_1 = \lambda + \mu_1 + 2\alpha,$$

$$v_2 = \lambda + \mu_2 + 2\alpha,$$

$$v_3 = \lambda + \mu_3 + 2\alpha,$$

$$v_{12} = \lambda + \mu_1 + \mu_2 + 2\alpha,$$

$$v_{23} = \lambda + \mu_2 + \mu_3 + 2\alpha,$$

$$v_{31} = \lambda + \mu_3 + \mu_1 + 2\alpha,$$

$$v_{123} = \lambda + \mu_1 + \mu_2 + \mu_3 + 3\alpha,$$

$$w_{12} = \mu_1 + \mu_2/2,$$

$$w_{21} = \mu_2 + \mu_1/2,$$

$$w_{23} = \mu_2 + \mu_3/2,$$

$$w_{32} = \mu_3 + \mu_2/2,$$

$$w_{31} = \mu_3 + \mu_1/2,$$

$$w_{13} = \mu_1 + \mu_3/2, \tag{2.76}$$

and

$$\mu' = \mu_1 + \mu_2 + \mu_3. \tag{2.77}$$

The generator $\mathbf{Q}$, partitioned as shown in (2.74), indicates that the continuous time Markov chain is a QBD process. Thus it is easy to apply the results developed by Neuts [71]. Specifically, the process is positive recurrent if and only if $\pi\mathbf{A_2e} > \pi\mathbf{A_0e}$, where π is the stationary probability vector of $\mathbf{A} = \mathbf{A_0} + \mathbf{A_1} + \mathbf{A_2}$ and the transpose of $\mathbf{e}$ is $(1, 1, \cdots, 1)$ with a proper size.

Equivalently, the system is stable if and only if the traffic intensity $\rho < 1$, where

$\rho = \pi \mathbf{A_0 e} / \pi \mathbf{A_2 e}$. For this model, it is easy to verify that

$$\rho = \frac{\lambda \pi_7}{(\mu' + 3\alpha)(\pi_1 + \pi_2 + \pi_3) + 3\alpha(\pi_4 + \pi_5 + \pi_6)}. \tag{2.78}$$

Solving the equations $\pi \mathbf{A} = \mathbf{0}$ and $\pi \mathbf{e} = 1$, we obtain

$$\pi_1 + \pi_2 + \pi_3 = \pi_4 + \pi_5 + \pi_6 = \pi_7 = \frac{1}{3}. \tag{2.79}$$

Hence

$$\rho = \frac{\lambda}{\mu' + 6\alpha}. \tag{2.80}$$

If $\mu_1 = \mu_2 = \mu_3 = \mu$, then it is identical to the traffic intensity of $M/MVE/3(\infty)$ with equal service rate given by (2.21) for $c = 3$.

2.3.2. Stationary Probability Vector

Because of the boundary conditions, the stationary probability vector $\mathbf{x} = [\mathbf{x_0}, \mathbf{x_1}, \cdots]$ is a modified matrix-geometric invariant vector with $\mathbf{x_k} = \mathbf{x_1 R^{k-1}}$, $k = 1, 2, \cdots$, where the rate matrix $\mathbf{R}$ is the minimal nonnegative solution of the matrix-quadratic equation

$$\mathbf{R^2 A_2} + \mathbf{R A_1} + \mathbf{A_0} = \mathbf{0}, \tag{2.81}$$

(e.g., see Neuts [71, 72]). To find $(\mathbf{x_0}, \mathbf{x_1})$, we use the boundary conditions

$$\begin{aligned} \mathbf{x_0 A_{01}} + \mathbf{x_1 A_{12}} &= \mathbf{0}, \\ \mathbf{x_0 A_{00}} + \mathbf{x_1 A_{11}} + \mathbf{x_2 A_2} &= \mathbf{0}, \end{aligned} \tag{2.82}$$

and the normalization condition

$$\mathbf{x_0} + \mathbf{x_1 (I - R)^{-1} e} = 1. \tag{2.83}$$

Using (2.82) and (2.83), we solve the following system of linear equations for $(\mathbf{x_0}, \mathbf{x_1})$:

$$(\mathbf{x_0}, \mathbf{x_1}) \begin{bmatrix} \mathbf{A_{00}} & 1 \\ \mathbf{A_{11}} + \mathbf{R}\mathbf{A_2} & (\mathbf{I} - \mathbf{R})^{-1}\mathbf{e} \end{bmatrix} = (\mathbf{0}, 1), \tag{2.84}$$

where $\mathbf{0}$ is a zerovector.

The rate matrix $\mathbf{R}$ is generally found by a number of iterative methods (e.g., see Ramaswami [90]). However, the generator $\mathbf{Q}$ defined by (2.74) has a very special structure. But the structure does not enable us to use the results obtained by Gillent and Latouche [27], and Ramaswami and Latouche [89] in the same manner as Kao and Lin [38], because the matrix $\mathbf{A_0}$ is given by $\mathbf{A_0} = \mathbf{p} \cdot \mathbf{m}$ but the matrix $\mathbf{A_2}$ is not given by $\mathbf{A_2} = \mathbf{q} \cdot \mathbf{n}$, where $\mathbf{p}$ and $\mathbf{q}$ are column vectors, and $\mathbf{m}$ and $\mathbf{n}$ are row vectors. Using the results given by Adan, Wessels and Zijm [1], the matrix $\mathbf{R}$ can be determined explicitly and many interesting features about the model can be explored conveniently.

We have $\mathbf{A_0} = \mathbf{e_7} \cdot \omega$, where $\mathbf{e_7}$ is a column vector of zeros except its last (the seventh) element is a one. We write the equation (2.81) as follows:

$$\mathbf{R} = -\mathbf{A_0}(\mathbf{A_1} + \mathbf{R}\mathbf{A_2})^{-1}. \tag{2.85}$$

Since $\mathbf{A_0} = \mathbf{e_7} \cdot \omega$, (2.85) implies that all rows of $\mathbf{R}$ are zero except the last—an observation to be expected from Theorem 1.3.4 of Neuts [71]. Denote the last row of $\mathbf{R}$ by $\mathbf{r} = \{r_1, \cdots, r_7\}$. It is easy to verify that

$$x_{ki} = \mathbf{x_1}(r_7)^{k-2} r_i, \quad 1 \leq i \leq 7, k \geq 2. \tag{2.86}$$

Let η be the spectral radius of $\mathbf{R}$, then

$$\eta = \mathbf{x_1}\mathbf{R^{k+1}}\mathbf{e}/\mathbf{x_1}\mathbf{R^k}\mathbf{e} = r_7, \quad k \geq 1, \tag{2.87}$$

(e.g.,see Neuts [73]).

This system can be represented by a continuous time Markov process whose state space S consists of the vectors (s_1, s_2, s_3) where s_i is the length of queue $i, i = 1, 2, 3$. Let V_l be the set of states $(s_1, s_2, s_3) \in S$ satisfying $s_1 + s_2 + s_3 = l$ and $P(V_l)$ be the steady state probability for the set V_l. Similarly to the equal case described in Section 2.2, the steady state probabilities of $M/MVE/3(\infty)$ with different service rate are given as follows:

$$p_{n+3} = (1/\zeta)^3 p_n, \quad n \geq 3, \tag{2.88}$$

where

$$\zeta = [\mu_1 + \mu_2 + \mu_3 + 3\alpha + \{(\mu_1 + \mu_2 + \mu_3 + 3\alpha)^2 + 12\alpha\lambda\}^{\frac{1}{2}}]/(2\lambda). \tag{2.89}$$

Therefore, by balancing the flow between the sets V_l and V_{l+3} it follows that

$$P(V_{l+3}) = (1/\zeta)^3 P(V_l), \quad l \geq 3. \tag{2.90}$$

On the other hand, (2.86) implies that if $\max(s_1, s_2, s_3) > 1$

$$P(s_1 + 1, s_2 + 1, s_3 + 1) = r_7 P(s_1, s_2, s_3), \tag{2.91}$$

so it follows that $l > 3$

$$P(V_{l+3}) = r_7 P(V_l). \tag{2.92}$$

Combining (2.90) and (2.92) yields

$$r_7 = (1/\zeta)^3. \tag{2.93}$$

The remaining components of $\mathbf{r}$ are obtained from the equation (2.81). By insertion of the special forms of $\mathbf{R}$ and $\mathbf{A_0}$ (2.81) is simplified to

$$\omega + \mathbf{r}(\mathbf{A_1} + r_7 \mathbf{A_2}) = \mathbf{0}. \tag{2.94}$$

Finally, substitution of (2.93) into (2.94) leads to

$$\mathbf{r} = -\omega\{\mathbf{A_1} + (1/\zeta)^3\mathbf{A_2}\}^{-1}. \tag{2.95}$$

These results are summarized in the following theorem:

Theorem 2.1. *The rate matrix* **R** *which is the minimal nonnegative solution of (2.81) is given by*

$$\mathbf{R} = \mathbf{e_7} \cdot \mathbf{r}, \tag{2.96}$$

where

$$\mathbf{r} = -\omega\{\mathbf{A_1} + (1/\zeta)^3\mathbf{A_2}\}^{-1}.$$

The rate matrix **R** is given in an explicit form by the above theorem and can be found without having to use any iterative method. Hence, by solving equations (2.82) and (2.83), we can find $(\mathbf{x_0}, \mathbf{x_1})$, and also $\mathbf{x_k}(k \geq 2)$ from (2.86). Consequently, by using these stationary probabilities and **R**, we can calculate not only the mean queue length but also the mean waiting time in the same manner as Kao and Lin [38].

Further though we have developed the solution for jockeying problem for the case of three servers ($c = 3$), the generator **Q** of the continuous time Markov chain for the general case with $c > 3$ can be similarly constructed once the state space is defined, and our results follow accordingly.

2.4. Conclusion

We considered a multiserver queueing system with correlated service times whose distribution is the multivariate exponential distribution of Marshall and Olkin in this chapter. Both jockeying and no jockeying problems for this model were analyzed. For the no jockeying problem, we obtained the steady state probabilities, the queueing and waiting time distribution and showed that Little's formula is valid for their mean, by a traditional generating function approach. On the other side, for the jockeying problem, we showed how to calculate explicitly the rate matrix R using the results derived in the no jockeying problem by a matrix-geometric approach.

Chapter 3.

TANDEM QUEUEING SYSTEMS WITH CORRELATED SERVICE TIMES

3.1. Introduction

This chapter discusses a correlated two or three stage tandem queueing system with a single station in which the service times of each station have the bivariate or the multivariate exponential distribution of Marshall and Olkin discussed in Chapter 1.

In many tandem queueing models having been investigated by several researchers, customers on arrival must wait when the station in which service is performed in the usual order is on service. Customers are processed through the ordered sequence of stages. We call this model as an ordinary tandem queueing system in this thesis. However, in practical situation, there are some cases where empty stations can be used for customers regardless of the ordered sequence of stages in order to increase the performance of the system. We call this model as a commutative tandem queueing system. Commutative tandem queueing systems can be explained precisely as follows: When the station in which service is performed in the usual order is busy, a customer can enter the other station if it is empty, and after completion of this service, he can enter any other empty station.

Nishida et al. [76] first considered a two-stage commutative tandem queueing system with no intermediate buffers. The mean queue length in the queue was obtained in the case of infinite possible queue ahead of the first stage. Nishida and Hiramatsu [77] also derived the mean queue length in the case of finite possible queue in front of the first stage. Furthermore, Nishida and Hiramatsu [78] analyzed a two-stage commutative tandem queueing system in which the service times of each station have the bivariate exponential distribution of Marshall and Olkin, derived the mean queue length and showed that it decreases when the correlation parameter increases as some numerical results.

In Section 3.2, we consider a two-stage ordinary tandem queueing system with no intermediate buffers. The successive customers arrive according to a homogeneous Poisson process. An unlimited queue is allowed in front of the first stage. The service times of two stages are assumed to follow the bivariate exponential distribution of Marshall and Olkin. For this model, we calculate the mean number of customers and show analytically that it decreases when the correlation parameter increases.

In Section 3.3, we consider a three-stage commutative tandem queueing system in which the service times of each stage are exponentially distributed and the assumption of waiting spaces and arrival process are as the same as treated in Section 3.2. For this system, we get the critical input rate, i.e., the throughput rate of the system using a matrix-geometric approach, and compare the throughput of this system with that of a three-stage ordinary tandem queueing system with correlated service times whose distribution is the multivariate exponential distribution of Marshall and Olkin.

In Section 3.4, first we consider a correlated two-server parallel queueing system discussed in Section 2.2 and secondly a correlated two-stage ordinary tandem queueing system which will be discussed in Section 3.2 for both cases where their service distribution is the bivariate exponential distribution of Marshall and Olkin. We show that the mean number of customers in the system is greater in the parallel case but lower in the tandem case than under the assumption of mutually independent service times when the system can finish 2μ customers' service per unit time on the average. In Section 3.5, we have a summary of this chapter.

3.2. Two-Stage Tandem Queueing System with Correlated Service Times

3.2.1. Model

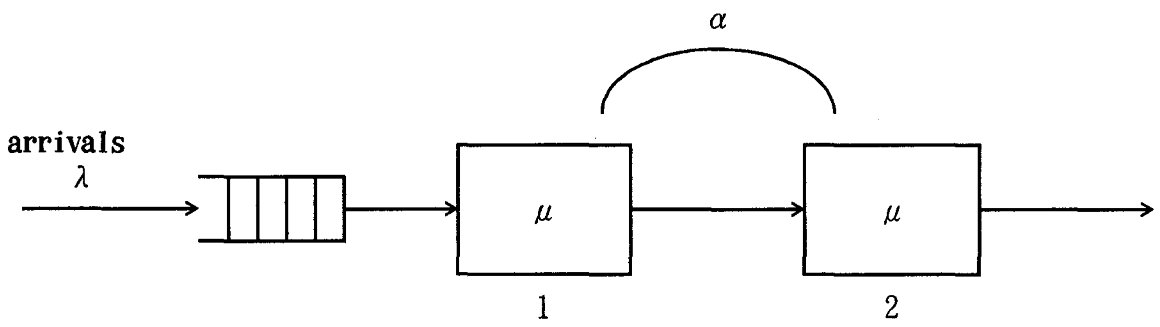

Fig. 3-1. Correlated two-stage tandem queueing system.

Fig. 3-1 shows a tandem queueing system with two stages numbered 1, 2. Stage $i(i = 1, 2)$ has a single server and the service times are assumed to follow the bivariate exponential distribution of Marshall and Olkin.

Namely, denoting by $X_i (i = 1, 2)$ the service times of two stages, its joint distribution is defined as follows:

$$Pr[X_1 > x_1, X_2 > x_2] = \exp[-\mu(x_1 + x_2) - \alpha \max(x_1, x_2)]. \tag{3.1}$$

The correlation coefficient of X_1 and X_2 is

$$\rho[X_1, X_2] = \alpha/(2\mu + \alpha). \tag{3.2}$$

An unlimited queue is allowed in front of stage 1. The successive customers arrive according to a Poisson process with rate $\lambda > 0$. Customers require service from all stages in the order $1, 2$. No queues are permitted between stage 1 and stage 2. This restriction results in the blocking of stage 1 whenever a customer having completed his service in stage 1 cannot move into stage 2 due to the presence of another customer there. It is also assumed that customers can transfer between stages instantaneously. Service to a customer at any stage, once initiated, is completed without interruptions. A queue discipline is first-come first-served.

3.2.2. Calculation of Mean Number of Customers

At any time t the customer who has not yet finished service in stage 1 is called a type 1 customer while one who has finished service in stage 1 but not yet in stage 2 is called a type 2 customer.

Let $Y_i(t)(i = 1, 2)$ be the number of type i customers in the system at any time t. Let also

$$p_{n,m} = Pr[Y_1(t) = n, Y_2(t) = m], \quad n \geq 0,\ m = 0, 1, 2. \tag{3.3}$$

As the Markov process governing the states of one system irreducible, the limiting value

$$\lim_{t\to\infty} p_{n,m}(t) = p_{n,m} \tag{3.4}$$

always exists. Thus connecting, as usual, $p_{n,m}(t+dt)$ with $p_{n,m}(t)$ and assuming that the system has reached statistical equilibrium we obtain the following system of difference equations.

$$\begin{aligned}
\lambda p_{0,0} &= (\mu+\alpha)p_{0,1}, \\
(\lambda+\mu+\alpha)p_{0,1} &= (\mu+\alpha)p_{0,2} + (\mu+\alpha)p_{1,0} + \alpha p_{1,1}, \\
(\lambda+\mu+\alpha)p_{0,2} &= \mu p_{1,1}, \\
(\lambda+\mu+\alpha)p_{n,0} &= \lambda p_{n-1,0} + \mu p_{n,1}, && n \geq 0, \\
(\lambda+2\mu+\alpha)p_{n,1} &= \lambda p_{n-1,1} + (\mu+\alpha)p_{n,2} + (\mu+\alpha)p_{n+1,0} + \alpha p_{n+1,1}, && n \geq 0, \\
(\lambda+\mu+\alpha)p_{n,2} &= \lambda p_{n-1,2} + \mu p_{n+1,1}, && n \geq 0.
\end{aligned} \tag{3.5}$$

Introducing now the generating functions

$$F_{m+1}(z) = \sum_{n=0}^{\infty} p_{n,m} z^{n+m}, \qquad m = 0, 1, 2; |z| \leq 1, \tag{3.6}$$

we obtain from (3.5)

$$\begin{aligned}
& z\{\rho(1-z)+1+\theta\}F_1(z) - F_2(z) = (1+\theta)zp_{0,0} + \theta z p_{0,1}, \\
& (1+\theta)zF_1(z) - \{\rho z(1-z) + (2+\theta)z - \theta\}F_2(z) + (1+\theta)F_3(z) \\
& \qquad = (1+\theta)zp_{0,0} + z(\theta - z)p_{0,1}, \\
& F_2(z) - \{\rho(1-z)+1+\theta\}F_3(z) = zp_{0,1},
\end{aligned} \tag{3.7}$$

where

$$\rho = \frac{\lambda}{\mu} \quad \text{and} \quad \theta = \frac{\alpha}{\mu}. \tag{3.8}$$

From (3.7), we have

$$F_i(z) = \frac{H_i(z)}{\Delta(z)}, \quad i = 1, 2, 3, \tag{3.9}$$

where

$$\Delta(z) = \begin{vmatrix} z\{\rho(1-z)+1+\theta\} & -1 & 0 \\ (1+\theta)z & -\{\rho z(1-z)+(2+\theta)z-\theta\} & 1+\theta \\ 0 & 1 & -\{\rho(1-z)+1+\theta\} \end{vmatrix}$$

$$= z(1-z)\{\rho(1-z)+1+\theta\}\{\rho^2 z(1-z)+\rho(3+2\theta)z-(1+\theta)(2+\theta)-\rho\theta\}, \tag{3.10}$$

and $H_i(z)(i = 1, 2, 3)$ is the determinant obtained from $\Delta(z)$ by replacing the i-th column by a column vector

$$\begin{bmatrix} (1+\theta)zp_{0,0}+\theta zp_{0,1} \\ (1+\theta)zp_{0,0}+z(\theta-z)p_{0,1} \\ zp_{0,1} \end{bmatrix}. \tag{3.11}$$

From (3.10), it is easily verified that

$$\left[\frac{H_1(z)}{1-z}\right]_{z=1} > 0. \tag{3.12}$$

Hence, since

$$0 < F_1(z) \leq 1, \tag{3.13}$$

we obtain

$$\left[\frac{\Delta(z)}{1-z}\right]_{z=1} > 0, \tag{3.14}$$

namely

$$\rho < \frac{(1+\theta)(2+\theta)}{3+\theta}. \tag{3.15}$$

If $\alpha = 0$, the service times are mutually independent, hence by letting $\theta \to 0$ in (3.15),

$$\rho < \frac{2}{3}, \tag{3.16}$$

which agrees with the steady state condition of the ordinary tandem queueing system [86]. Therefore, it follows that compared with the ordinary case, the maximum utilization of this system is increased when $\alpha > 0$.

Returning to (3.7), we get the generating functions $F_i(z)(i = 1,2,3)$ by elementary but a little troublesome calculation as follows.

$$\begin{aligned}
F_1(z) &= z(1-z)[(1+\theta)\{\rho^2 z(1-z) + \rho(3+2\theta)z - (1+\theta)(\rho+2+\theta)\}p_{0,0} \\
&\quad + \{\rho^2\theta z(1-z) + \rho(1+\theta)(1+2\theta)z - (1+\theta)^3 - \rho\theta(1+\theta)\}p_{0,1}]/\Delta(z), \\
F_2(z) &= -z^2(1-z)\{\rho(1-z)+1+\theta\}\{\rho(1+\theta)p_{0,0} + (1+\theta+\rho\theta-\rho z)p_{0,1}\}/\Delta(z), \\
F_3(z) &= -z^2(1-z)[\rho(1+\theta)p_{0,0} + \{\rho^2 z(1-z) + 2\rho(1+\theta)z - (1+\theta)^2\}p_{0,1}]/\Delta(z),
\end{aligned} \tag{3.17}$$

where $\Delta(z)$ is given by (3.10).

To determine unknowns $p_{0,0}$ and $p_{0,1}$, we may use the normalization condition

$$\sum_{m=0}^{2} F_{m+1}(1) = 1, \tag{3.18}$$

hence

$$(1+\theta)(2+\theta)p_{0,0} + (1+\theta)^2 p_{0,1} = (1+\theta)(2+\theta) - \rho(3+\theta). \tag{3.19}$$

From (3.19) and the first equation in (3.5), we find

$$p_{0,0} = \frac{(1+\theta)(2+\theta) - \rho(3+\theta)}{(1+\theta)(\rho+2+\theta)}. \tag{3.20}$$

If $\theta = 0$, (3.20) yields

$$\lim_{\theta\to 0} p_{0,0} = \frac{2-3\rho}{2+\rho}, \tag{3.21}$$

which coincides with that of the ordinary case [86].

Now we shall derive the mean number of customers in this system. In order to determine this, we must obtain $F_i'(1)(i = 1,2,3)$ explicitly. Differentiating $F_i(z)$ and letting $z = 1$, we can solve $F_i'(1)$ by some calculations and the mean number of customers L is given by

$$\begin{aligned} L &= \sum_{n=0}^{\infty}\sum_{m=0}^{2}(n+m)p_{n,m} \\ &= \sum_{i=1}^{3} F_i'(1) \\ &= \frac{2\rho(1+\theta)^2(2+\theta)^2 + \rho^2\theta(1+\theta)(2+\theta) - \rho^3(4+3\theta+\theta^2)}{(1+\theta)^2(\rho+2+\theta)\{(1+\theta)(2+\theta)-\rho(3+\theta)\}}. \end{aligned} \tag{3.22}$$

If $\theta = 0$, (3.22) is identical with

$$\lim_{\theta\to 0} L = \frac{4\rho(2-\rho^2)}{(2+\rho)(2-3\rho)}, \tag{3.23}$$

in the ordinary case [86].

3.2.3. Monotonicity of Mean Number of Customers

Let $L = L(\theta)$ be the mean number of customers. We then have the following theorem.

Theorem 3.1. *L is strictly decreasing with respect to θ.*

Proof. Differentiating L in (3.22) with respect to θ and considering the steady state condition (3.15), we have

$$\begin{aligned} L'(\theta) = & \frac{-\rho}{(1+\theta)^3(\rho+2+\theta)^2\{(1+\theta)(2+\theta)-\rho(3+\theta)\}^2} \\ & \cdot\{2(1+\theta)^3(2+\theta)^4 + 2\rho(1+\theta)^2(2+\theta)^2(1+3\theta+\theta^2) \\ & -\rho^2(1+\theta)(2+\theta)(14+11\theta+2\theta^2+\theta^3) + 2\rho^3(17+20\theta+9\theta^2+2\theta^3) \\ & +\rho^4(19+15\theta+5\theta^2+\theta^3)\} \end{aligned}$$

$$< \frac{-\rho^2}{(1+\theta)^3(3+\theta)(\rho+2+\theta)^2\{(1+\theta)(2+\theta)-\rho(3+\theta)\}^2}$$

$$\cdot \{(1+\theta)^2(2+\theta)^2(28+51\theta+26\theta^2+3\theta^3)+2\rho^2(3+\theta)(17+20\theta+9\theta^2+2\theta^3)$$

$$+\rho^3(3+\theta)(19+15\theta+5\theta^2+\theta^3)\} < 0. \quad \square \tag{3.24}$$

Since we put $\theta = \alpha/\mu$, we find that $L(\alpha)$ is strictly decreasing with respect to the correlation parameter α when μ is constant. Consequently, from (3.2) positive correlation is found to decrease the mean number of customers in this system.

3.3. Three-Stage Commutative Tandem Queueing System with Correlated Service Times

3.3.1. Model

Fig. 3-2 shows a tandem queueing system consisting of three single-server stations with no intermediate buffers, where customers arrive at the queue, which is assumed to be infinite, according to a homogeneous Poisson process with rate $\lambda > 0$. Each customer must enter the first station if it is free, and he joins the queue if it is not free. After completion of service in the first station and when the second station is busy, he can enter the third station if it is free, and after completion of this service, he can enter the second station. If a customer has already completed service of three stations, then he departs the system. If a customer has already completed the first service and both the second and third station are busy, he has to stay in the first station, that is to say, the first station is blocked. If he has not completed all service of three stations and the station in which he receives the third service is not free, he has to stay there, that is to say, this station is blocked, and when it has completed service, he can enter that station.

The service time for station i $(i = 1, 2, 3)$ is exponentially distributed with parameter μ_i. It is assumed that customers can transfer between stations instantaneously. Service to a customer at any stage, once initiated, is completed without interruptions. The queueing discipline is first-come first-served.

The queueing model under consideration can be studied as a continuous time Markov chain with state-space $\{(0,j) : 1 \leq j \leq m_0\} \cup \{(i,j) : i \geq 0, 1 \leq j \leq m\}$ The state $(0,j)$ denotes that the system is in the boundary state and m_0 denotes the number of the boundary states. In the state (i,j), i denotes the number of customers in the queue, whereas j denotes the state of the network consisting of three stations and m denotes the number of states in the network.

The states of the network are described by the vector:

$$(s_1, s_2, s_3) \tag{3.25}$$

where s_i $(i = 1, 2, 3)$ can take any value from 0 to 4 with

$$s_i = \begin{cases} 0 & \text{-the } i\text{-th station is idle,} \\ 1 & \text{-the } i\text{-th station is in the first service,} \\ 2 & \text{-the } i\text{-th station is in the second service,} \\ 3 & \text{-the } i\text{-th station is in the third service,} \\ 4 & \text{-the } i\text{-th station is blocked.} \end{cases} \tag{3.26}$$

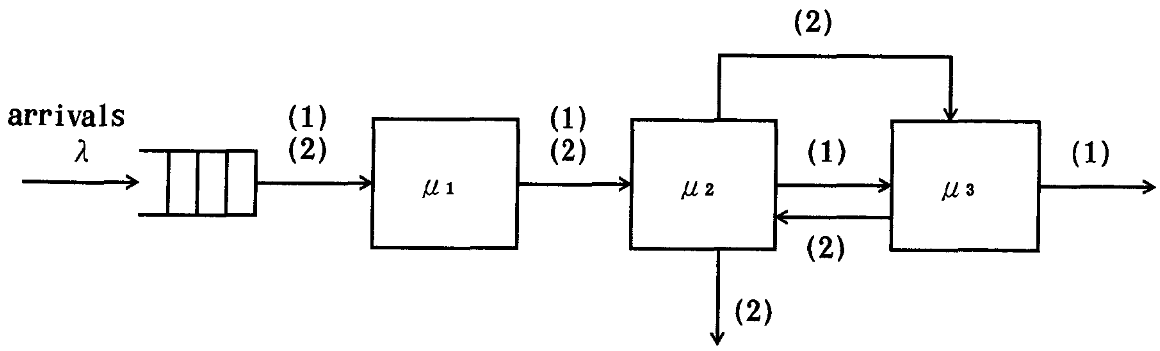

Fig. 3-2. Correlated three-stage commutative tandem queueing system.

Then the boundary states are

$$(0,0,0),(0,2,0),(0,0,2),(0,3,0),(0,0,3),(0,2,2),(0,3,3),$$

$$(0,2,3),(0,3,2),(0,2,4),(0,4,2),(0,3,4),(0,4,3)), \qquad (3.27)$$

whereas the states of the network are

$$(1,0,0),(1,2,0),(1,0,2),(1,3,0),(1,0,3),(1,2,2),(1,3,3),$$

$$(1,2,3),(1,3,2),(1,2,4),(1,4,2),(1,3,4),(1,4,3),(4,2,2),$$

$$(4,3,3),(4,2,3),(4,3,2),(4,2,4),(4,4,2),(4,3,4),(4,4,3). \qquad (3.28)$$

By ordering the states as described above, the infinitestimal generator of the continuous time Markov chain has the following block partitioned structure:

$$\mathbf{Q} = \begin{bmatrix} \mathbf{A_{01}} & \mathbf{A_{03}} & & & & & & \\ \mathbf{A_{02}} & \mathbf{A_1} & \mathbf{A_0} & & & & & \\ & \mathbf{A_2} & \mathbf{A_1} & \mathbf{A_0} & & & & \\ & & \mathbf{A_2} & \mathbf{A_1} & \mathbf{A_0} & & & \\ & & & \mathbf{A_2} & \mathbf{A_1} & \mathbf{A_0} & & \\ & & & & \mathbf{A_2} & \mathbf{A_1} & & \\ & & & & & \vdots & \ddots & \end{bmatrix}, \qquad (3.29)$$

where all the unmarked entries are zeros.

The submatrices are defined as below. The dimensionality of $\mathbf{A_{01}}$ is 13×13, $\mathbf{A_{02}}$ is 21×13, $\mathbf{A_{03}}$ is 13×21, $\mathbf{A_0}$, $\mathbf{A_1}$ and $\mathbf{A_2}$ are 21×21. More specifically,

$$\mathbf{A_{01}} = \text{diag}\,(-\lambda, -u_2, -u_3, -u_2, -u_3, -u_{23}, -u_{23},$$

$$-u_{23}, -u_{23}, -u_2, -u_3, -u_2, -u_3)$$

$$+\mu_2\,(\mathbf{e_6} \cdot \mathbf{e_{11}'} + \mathbf{e_7} \cdot \mathbf{e_5'} + \mathbf{e_8} \cdot \mathbf{e_{13}'}$$

$$+ \mathbf{e_9} \cdot \mathbf{e'_3} + \mathbf{e_{10}} \cdot \mathbf{e'_7} + \mathbf{e_{12}} \cdot \mathbf{e'_4})$$

$$+ \mu_3 \, (\mathbf{e_6} \cdot \mathbf{e'_{10}} + \mathbf{e_7} \cdot \mathbf{e'_4} + \mathbf{e_8} \cdot \mathbf{e'_2}$$

$$+ \mathbf{e_9} \cdot \mathbf{e'_{12}} + \mathbf{e_{11}} \cdot \mathbf{e'_7} + \mathbf{e_{13}} \cdot \mathbf{e'_5}),$$

$$\mathbf{A_{02}} = \mu_1 \, (\mathbf{e_1} \cdot \mathbf{e'_2} + \mathbf{e_2} \cdot \mathbf{e'_6} + \mathbf{e_3} \cdot \mathbf{e'_6}$$

$$+ \mathbf{e_4} \cdot \mathbf{e'_9} + \mathbf{e_5} \cdot \mathbf{e'_8})$$

$$+ \mu_2 \, (\mathbf{e_{15}} \cdot \mathbf{e'_8} + \mathbf{e_{17}} \cdot \mathbf{e'_6} + \mathbf{e_{20}} \cdot \mathbf{e'_9})$$

$$+ \mu_3 \, (\mathbf{e_{15}} \cdot \mathbf{e'_9} + \mathbf{e_{16}} \cdot \mathbf{e'_6} + \mathbf{e_{21}} \cdot \mathbf{e'_8}),$$

$$\mathbf{A_{03}} = (\, \lambda \, \mathbf{I_{13}} \quad \mathbf{O_x} \,) \,,$$

$$\mathbf{A_0} = \lambda \, \mathbf{I},$$

$$\mathbf{A_1} = \mathrm{diag}\,(-u_1, -u_{12}, -u_{31}, -u_{12}, -u_{31}, -u_{123}, -u_{123},$$

$$-\, u_{123}, -u_{123}, -u_{12}, -u_{31}, -u_{12}, -u_{31}, -u_{23},$$

$$-\, u_{23}, -u_{23}, -u_{23}, -u_2, -u_3, -u_2, -u_3)$$

$$+ \mu_1 \, \sum_{\mathbf{i=1}}^{\mathbf{8}} \mathbf{e_{i+5}} \cdot \mathbf{e'_{i+13}}$$

$$+ \mu_2 \, (\mathbf{e_2} \cdot \mathbf{e'_5} + \mathbf{e_4} \cdot \mathbf{e'_1} + \mathbf{e_6} \cdot \mathbf{e'_{11}}$$

$$+ \mathbf{e_7} \cdot \mathbf{e'_5} + \mathbf{e_8} \cdot \mathbf{e'_{13}} + \mathbf{e_9} \cdot \mathbf{e'_3}$$

$$+ \mathbf{e_{10}} \cdot \mathbf{e'_7} + \mathbf{e_{12}} \cdot \mathbf{e'_4} + \mathbf{e_{14}} \cdot \mathbf{e'_{19}}$$

$$+\mathbf{e_{16}}\cdot\mathbf{e}_{21}'+\mathbf{e_{18}}\cdot\mathbf{e}_{15}')$$

$$+\mu_3\,(\mathbf{e_3}\cdot\mathbf{e}_4'+\mathbf{e_5}\cdot\mathbf{e}_1'+\mathbf{e_6}\cdot\mathbf{e}_{10}'$$

$$+\mathbf{e_7}\cdot\mathbf{e}_4'+\mathbf{e_8}\cdot\mathbf{e}_2'+\mathbf{e_9}\cdot\mathbf{e}_{12}'$$

$$+\mathbf{e_{11}}\cdot\mathbf{e}_7'+\mathbf{e_{13}}\cdot\mathbf{e}_5'+\mathbf{e_{14}}\cdot\mathbf{e}_{18}'$$

$$+\mathbf{e_{17}}\cdot\mathbf{e}_{20}'+\mathbf{e_{19}}\cdot\mathbf{e}_{15}'),$$

$$\mathbf{A_2}=(\,\mathbf{A_{02}}\quad\mathbf{O_y}\,), \tag{3.30}$$

where

$$u_1=\lambda+\mu_1,$$

$$u_2=\lambda+\mu_2,$$

$$u_3=\lambda+\mu_3,$$

$$u_{12}=\lambda+\mu_1+\mu_2,$$

$$u_{23}=\lambda+\mu_2+\mu_3,$$

$$u_{31}=\lambda+\mu_3+\mu_1,$$

$$u_{123}=\lambda+\mu_1+\mu_2+\mu_3, \tag{3.31}$$

and $\mathbf{I_{13}}$ is (13×13) unit matrix, $\mathbf{O_x}$ is (13×8) zero matrix, $\mathbf{O_y}$ is (21×8) zero matrix, $\mathbf{e_i}$ is a column vector of zeros except the i-th element is a one and $\mathbf{e}_i'$ is its transpose.

3.3.2. Calculation of Throughput

Let $\mathbf{p}$ equal the steady state probabilities of the network, assuming that the queue is never empty, which has elements $p(j)$ $(1\le j\le 21)$. We can determine $\mathbf{p}$, by solving the system

$$\mathbf{pA} = \mathbf{0}, \qquad \mathbf{pe} = 1 \tag{3.32}$$

where $\mathbf{A}$ is the conservative matrix given by

$$\mathbf{A} = \mathbf{A_0} + \mathbf{A_1} + \mathbf{A_2} \tag{3.33}$$

and $\mathbf{e}$ is (21×1) column vector with all elements equal 1. The equilibrium condition is given (see Neuts[71]) by

$$\mathbf{pA_0e} < \mathbf{pA_2e}. \tag{3.34}$$

From this relationship the critical input rate, λ^*, to the system can be determined. In the steady state, this critical input rate is identical to the maximum throughput rate of the system.

The conservative matrix of our model becomes

$$\begin{aligned}
\mathbf{A} =& \text{diag}\ (-\mu_1, \lambda - u_{12}, \lambda - u_{31}, \lambda - u_{12}, \lambda - u_{31}, \lambda - u_{123}, \lambda - u_{123}, \\
& \lambda - u_{123}, \lambda - u_{123}, \lambda - u_{12}, \lambda - u_{31}, \lambda - u_{12}, \lambda - u_{31}, \lambda - u_{23}, \\
& \lambda - u_{23}, \lambda - u_{23}, \lambda - u_{23}, -\mu_2, -\mu_3, -\mu_2, -\mu_3) \\
& + \mu_1\ (\mathbf{e_1} \cdot \mathbf{e_2'} + \mathbf{e_2} \cdot \mathbf{e_6'} + \mathbf{e_3} \cdot \mathbf{e_6'} \\
& + \mathbf{e_4} \cdot \mathbf{e_9'} + \mathbf{e_5} \cdot \mathbf{e_8'} + \sum_{\mathbf{i}=1}^{8} \mathbf{e_{i+5}} \cdot \mathbf{e_{i+13}'}) \\
& + \mu_2\ (\mathbf{e_2} \cdot \mathbf{e_5'} + \mathbf{e_4} \cdot \mathbf{e_1'} + \mathbf{e_6} \cdot \mathbf{e_{11}'} \\
& + \mathbf{e_7} \cdot \mathbf{e_5'} + \mathbf{e_8} \cdot \mathbf{e_{13}'} + \mathbf{e_9} \cdot \mathbf{e_3'} \\
& + \mathbf{e_{10}} \cdot \mathbf{e_7'} + \mathbf{e_{12}} \cdot \mathbf{e_4'} + \mathbf{e_{14}} \cdot \mathbf{e_{19}'} \\
& + \mathbf{e_{15}} \cdot \mathbf{e_8'} + \mathbf{e_{16}} \cdot \mathbf{e_{21}'} + \mathbf{e_{17}} \cdot \mathbf{e_6'}
\end{aligned}$$

$$+\mathbf{e}_{18}\cdot\mathbf{e}'_{15}+\mathbf{e}_{20}\cdot\mathbf{e}'_{9})$$

$$+\mu_3\,(\mathbf{e}_3\cdot\mathbf{e}'_4+\mathbf{e}_5\cdot\mathbf{e}'_1+\mathbf{e}_6\cdot\mathbf{e}'_{10}$$

$$+\mathbf{e}_7\cdot\mathbf{e}'_4+\mathbf{e}_8\cdot\mathbf{e}'_2+\mathbf{e}_9\cdot\mathbf{e}'_{12}$$

$$+\mathbf{e}_{11}\cdot\mathbf{e}'_7+\mathbf{e}_{13}\cdot\mathbf{e}'_5+\mathbf{e}_{14}\cdot\mathbf{e}'_{18}$$

$$+\mathbf{e}_{15}\cdot\mathbf{e}'_9+\mathbf{e}_{16}\cdot\mathbf{e}'_6+\mathbf{e}_{17}\cdot\mathbf{e}'_{20}$$

$$+\mathbf{e}_{19}\cdot\mathbf{e}'_{15}+\mathbf{e}_{21}\cdot\mathbf{e}'_8). \tag{3.35}$$

Putting $\mu_1=\mu_2=\mu_3=\mu$ and solving the system (3.32), we obtain:

$$p(1)=\frac{132}{1287},\ \ p(2)=\frac{109}{1287},\ \ p(4)=\frac{44}{1287},$$
$$p(5)=\frac{88}{1287},\ \ p(7)=\frac{24}{1287},\ \ p(15)=\frac{84}{1287},$$

$$\begin{aligned} p(3)=\ p(12)\ &=\ p(17)\ =\ \frac{32}{1287},\\ p(6)=\ p(18)\ &=\ p(19)\ =\ \frac{72}{1287},\\ p(10)=\ p(11)\ &=\ p(14)\ =\ \frac{36}{1287},\\ p(8)=\ p(21)\ &=\ \frac{86}{1287},\\ p(9)=\ p(20)\ &=\ \frac{64}{1287},\\ p(13)=\ p(16)\ &=\ \frac{43}{1287}. \end{aligned} \tag{3.36}$$

The equilibrium condition (3.34) in this case is given by

$$\lambda<\mu\,[\,p(1)+p(2)+p(3)+p(4)+p(5)+2p(15)+p(16)+p(17)+p(20)+p(21)\,]. \tag{3.37}$$

If $\mu=1$, we get

$$\lambda<\frac{798}{1287}=\lambda^*. \tag{3.38}$$

3.3.3. Throughput of Ordinary Tandem Queueing System with Correlated Service Times

We consider a correlated three-stage tandem queueing system in which services are performed in the usual order and their service distribution is the multivariate exponential distribution of Marshall and Olkin as follows:

$$Pr[X_1 > x_1, X_2 > x_2, X_3 > x_3] = \exp\left[-\mu_1 x_1 - \mu_2 x_2 - \mu_3 x_3 \right.$$
$$\left. - \alpha \max(x_1, x_2) - \alpha \max(x_2, x_3) - \alpha \max(x_3, x_1)\right]. \tag{3.39}$$

The remaining assumptions are just as same as the above three-stage commutative tandem queueing system. Using the results obtained by Papadopoulos and O'Kelly[83], the infinitesimal generator of the continuous time Markov chain has the following block partitioned structure:

$$\mathbf{Q} = \begin{bmatrix} \mathbf{A}_{01} & \mathbf{A}_0 & & & & \\ \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_0 & & & \\ & \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_0 & & \\ & & \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_0 & \\ & & & & \vdots & \ddots \end{bmatrix}. \tag{3.40}$$

The submatrices are defined as below. The dimensionalities of $\mathbf{A}_{01}$, $\mathbf{A}_0$, $\mathbf{A}_1$ and $\mathbf{A}_2$ are 8×8. More specifically,

$$\mathbf{A}_{01} = \begin{bmatrix} -\lambda & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -v_2 & w_2+\alpha & 0 & 0 & 0 & 0 & 0 \\ w_3+\alpha & 0 & -v_3 & 0 & 0 & 0 & 0 & 0 \\ 0 & w_3 & 0 & -v_{23} & w_2 & 0 & 0 & 0 \\ 0 & 0 & w_3+\alpha & 0 & -v_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & w_2+\alpha & 0 & -v_2 & 0 & 0 \\ 0 & 0 & 0 & \alpha & 0 & w_3 & -v_{23}-\alpha & w_2 \\ 0 & 0 & 0 & w_3+\alpha & 0 & 0 & 0 & -v_3 \end{bmatrix},$$

$$\mathbf{A}_0 = \lambda\, \mathbf{I},$$

$$\mathbf{A_1} = \begin{bmatrix} -v_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -v_{12}-\alpha & w_2 & 0 & 0 & 0 & 0 & 0 \\ w_3 & 0 & -v_{31}-\alpha & 0 & 0 & 0 & 0 & 0 \\ 0 & \mu_3 & \alpha & -v_{123}-\alpha & \mu_2 & 0 & 0 & 0 \\ 0 & 0 & w_3 & 0 & -v_{31}-\alpha & 0 & 0 & 0 \\ 0 & 0 & 0 & w_2+\alpha & 0 & -v_2 & 0 & 0 \\ 0 & 0 & 0 & \alpha & 0 & w_3 & -v_{23}-\alpha & w_2 \\ 0 & 0 & 0 & w_3+\alpha & 0 & 0 & 0 & -v_3 \end{bmatrix},$$

$$\mathbf{A_2} = \begin{bmatrix} 0 & w_1+\alpha & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha & 0 & w_1 & 0 & 0 \\ 0 & \alpha & 0 & w_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \alpha & \mu_1 & \alpha \\ 0 & 0 & 0 & \alpha & 0 & 0 & 0 & w_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{3.41}$$

where

$$v_1 = \lambda + \mu_1 + 2\alpha,$$

$$v_2 = \lambda + \mu_2 + 2\alpha,$$

$$v_3 = \lambda + \mu_3 + 2\alpha,$$

$$v_{12} = \lambda + \mu_1 + \mu_2 + 2\alpha,$$

$$v_{23} = \lambda + \mu_2 + \mu_3 + 2\alpha,$$

$$v_{31} = \lambda + \mu_3 + \mu_1 + 2\alpha,$$

$$v_{123} = \lambda + \mu_1 + \mu_2 + \mu_3 + 2\alpha,$$

$$w_1 = \mu_1 + \alpha,$$

$$w_2 = \mu_2 + \alpha,$$

$$w_3 = \mu_3 + \alpha. \tag{3.42}$$

Let $\mathbf{q}$ equal the steady state probabilities of the subnetwork, assuming that the queue is never empty, which elements $q(j)$ $(1 \leq j \leq 8)$. We can determine $\mathbf{q}$, by solving the system

$$\mathbf{qA} = \mathbf{0}, \quad \mathbf{qe} = 1 \tag{3.43}$$

where $\mathbf{A}$ is the conservative matrix given by

$$\mathbf{A} = \mathbf{A_0} + \mathbf{A_1} + \mathbf{A_2} \tag{3.44}$$

and $\mathbf{e}$ is (8×1) column vector with all elements equal 1. The equilibrium condition is given by

$$\mathbf{qA_0e} < \mathbf{qA_2e}. \tag{3.45}$$

The conservative matrix of this queueing system becomes

$$\mathbf{A} = \begin{bmatrix} -w_1-\alpha & w_1+\alpha & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -r_{12} & w_2 & \alpha & 0 & w_1 & 0 & 0 \\ w_3 & \alpha & -r_{31} & w_1 & 0 & 0 & 0 & 0 \\ 0 & \mu_3 & \alpha & -r_{123} & \mu_2 & \alpha & \mu_1 & \alpha \\ 0 & 0 & w_3 & \alpha & -r_{31} & 0 & 0 & w_1 \\ 0 & 0 & 0 & w_2+\alpha & 0 & -w_2-\alpha & 0 & 0 \\ 0 & 0 & 0 & \alpha & 0 & w_3 & -r_{23} & w_2 \\ 0 & 0 & 0 & w_3+\alpha & 0 & 0 & 0 & -w_3-\alpha \end{bmatrix}, \tag{3.46}$$

where

$$r_{12} = \mu_1 + \mu_2 + 3\alpha,$$

$$r_{23} = \mu_2 + \mu_3 + 3\alpha,$$

$$r_{31} = \mu_3 + \mu_1 + 3\alpha,$$

$$r_{123} = \mu_1 + \mu_2 + \mu_3 + 3\alpha. \tag{3.47}$$

Putting $\mu_1 = \mu_2 = \mu_3 = \mu$ and $\theta = \alpha/\mu$ and solving the system (3.43), we obtain:

$$
\begin{aligned}
q(1) &= \frac{(1+\theta)\ (3\theta^2+4\theta+2)}{(1+2\theta)\ (7\theta^2+9\theta+3)}\ q(4), \\
q(2) &= \frac{(1+\theta)\ (6\theta^2+12\theta+5)}{(2+3\theta)\ (7\theta^2+9\theta+3)}\ q(4), \\
q(3) &= \frac{3\theta^2+4\theta+2}{7\theta^2+9\theta+3}\ q(4), \\
q(4) &= \frac{(1+2\theta)\ (2+3\theta)\ (7\theta^2+9\theta+3)}{D}, \\
q(5) &= q(7) = \frac{1}{2+3\theta}\cdot q(4), \\
q(6) &= \frac{1}{1+2\theta}\cdot\left[\theta+\frac{1+\theta}{2+3\theta}+\frac{(1+\theta)^2\ (6\theta^2+12\theta+5)}{(2+3\theta)\ (7\theta^2+9\theta+2)}\right] q(4), \\
q(8) &= \frac{3\theta^2+4\theta+2}{(1+2\theta)\ (2+3\theta)}\ q(4),
\end{aligned}
\tag{3.48}
$$

where

$$
\begin{aligned}
D =&(1+\theta)\ (2+3\theta)\ (3\theta^2+4\theta+2)+(1+\theta)\ (1+2\theta)\ (6\theta^2+12\theta+5) \\
&+(1+2\theta)\ (2+3\theta)\ (3\theta^2+4\theta+2)+(1+2\theta)\ (2+3\theta)\ (7\theta^2+9\theta+3) \\
&+2(1+2\theta)\ (7\theta^2+9\theta+3)+\theta\ (2+3\theta)\ (7\theta^2+9\theta+3) \\
&+(1+\theta)\ (7\theta^2+9\theta+3)+(1+\theta)^2\ (6\theta^2+12\theta+5) \\
&+(3\theta^2+4\theta+2)\ (7\theta^2+9\theta+3).
\end{aligned}
\tag{3.49}
$$

The equilibrium condition (3.45) in this case is given by

$$
\lambda < \mu\ (1+2\theta)\ [q(1)+q(3)+q(6)+q(8)]+\mu\theta\ [q(2)+q(5)+q(7)]. \tag{3.50}
$$

If $\mu = 1$, we get

$$
\lambda < \frac{(1+2\theta)\ (81\theta^4+231\theta^3+250\theta^2+121\theta+22)}{129\theta^4+372\theta^3+412\theta^2+206\theta+39} = \lambda^*. \tag{3.51}
$$

When $\theta = 0$, that is, the service times of three stations are independent, the maximum throughput rate (3.51) yields

$$\lambda^* = \frac{22}{39}, \tag{3.52}$$

which coincides with the result given by Papadopoulos and O'Kelly[83].

3.3.4. Comparison of Throughput

Let the maximum throughput rate λ^* in (3.38), (3.51)and (3.52) denote $\lambda^*_{C,I}$,$\lambda^*_{O,R}$ and $\lambda^*_{O,I}$, respectively.

If

$$\theta = \theta_0 \approx 0.045, \tag{3.53}$$

then

$$\lambda^*_{C,I} = \lambda^*_{O,R}. \tag{3.54}$$

Consequently, since $\lambda^*_{O,R}$ is increasing function with respect to $\theta > 0$, we obtain the following result.

Theorem 3.2. *If $0 < \theta \le \theta_0$, then*

$$\lambda^*_{O,I} < \lambda^*_{O,R} \le \lambda^*_{C,I}. \tag{3.55}$$

And if $\theta > \theta_0$, then

$$\lambda^*_{O,I} < \lambda^*_{C,I} < \lambda^*_{O,R}. \tag{3.56}$$

The following explanation is derived from the above theorem. When the correlation parameter is small, a commutative tandem queueing model with independent service times improves the throughput of the system better than an ordinary tandem queueing model with correlated service times, by utilizing an empty station. When the correlation parameter is large, an ordinary tandem queueing model with correlated service times improves it better than a commutative tandem queueing model with independent service times, by the frequent occurrence of the event that two stations finish service at the same time.

3.4. Effect of Correlated Exponential Service Times on Queueing Systems

3.4.1. Two-Server Parallel Case

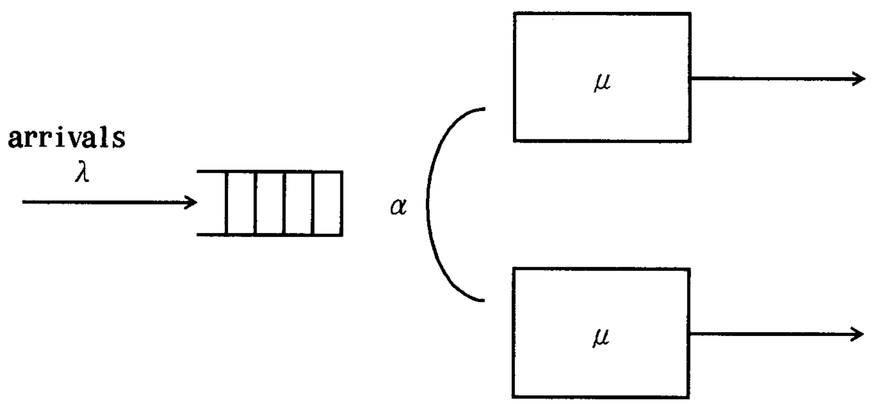

Fig. 3-3. The M/BVE/2(∞) queueing system.

Fig. 3-3 shows the $M/BVE/2(\infty)$ queueing system which is a special case of the queueing system described in Section 2.2. The service distribution of two servers is given by (3.1). We shall denote this system by $M/BVE(\mu, \mu, \alpha)/2(\infty)$. Letting c=2 in

(2.57) and (2.72), the mean number of customers in this system is

$$L_C = \frac{\lambda}{\mu+\alpha} + \frac{p_1}{(\zeta-1)^2} = \frac{\lambda}{\mu+\alpha} - \frac{\lambda\xi(2\mu+2\alpha-\lambda)}{(\zeta-1)^2(\mu+\alpha)\{(2\mu-\lambda)\xi+2\alpha\}}, \tag{3.57}$$

where

$$\xi = [2\mu+\alpha-\{(2\mu+\alpha)^2+4\alpha\lambda\}^{\frac{1}{2}}]/(2\lambda), \tag{3.58}$$

and

$$\zeta = [2\mu+\alpha+\{(2\mu+\alpha)^2+4\alpha\lambda\}^{\frac{1}{2}}]/(2\lambda). \tag{3.59}$$

From (2.21), the steady state condition is

$$\lambda < 2(\mu+\alpha). \tag{3.60}$$

On the other hand, we shall denote the ordinary two-server parallel queueing system by $M/M(\mu,\mu)/2(\infty)$. The mean number of customers in the $M/M(\mu,\mu)/2(\infty)$ system [28] is given by

$$L_I = \frac{2\rho}{1-\rho^2}, \tag{3.61}$$

where

$$\rho = \frac{\lambda}{2\mu}. \tag{3.62}$$

Since the event that two channels finish service at the same time occurs according to the rate α in addition to the occurrence of the event that only one channel finishes each service according to the rate μ independently, we can conjecture intuitively that

$$L_C < L_I \quad \text{for any } \rho. \tag{3.63}$$

And we find that when the event that two channels finish service at the same time occurs according to the rate α, then this system can finish 2α customers' service per unit time on the average. Consequently, it is natural that we should compare the mean number of customers of the $M/BVE(\mu-\alpha,\mu-\alpha,\alpha)/2(\infty)$ system with that of the $M/M(\mu,\mu)/2(\infty)$ system. Hence, replacing μ by $\mu-\alpha$ in (3.60) the steady state condition of the $M/BVE(\mu-\alpha,\mu-\alpha,\alpha)/2(\infty)$ system is

$$\lambda < 2\mu, \tag{3.64}$$

which coincides with one of the $M/M(\mu,\mu)/2(\infty)$ system. Then we have the following theorem:

Theorem 3.3. *The mean number of customers L_C of $M/BVE(\mu-\alpha,\mu-\alpha,\alpha)/2(\infty)$ is larger than L_I of $M/M(\mu,\mu)/2(\infty)$ if $0<\alpha<\mu$.*

Proof. Replacing μ by $\mu-\alpha$ in (3.57), and from (3.61) we have

$$L_I - L_C = \frac{2\rho\cdot\Delta L}{(1-\rho^2)(\zeta-1)^2\{(1-2\beta-\rho)\xi+2\beta\}}, \tag{3.65}$$

where

$$\begin{aligned}\Delta L &= \rho^2(\zeta-1)^2\{(1-2\beta-\rho)\xi+2\beta\}+(1-\rho)(1-\rho^2)\xi,\\ \xi &= [1-\beta-\{(1-\beta)^2+4\beta\rho\}^{\frac{1}{2}}]/(2\rho),\\ \zeta &= [1-\beta+\{(1-\beta)^2+4\beta\rho\}^{\frac{1}{2}}]/(2\rho),\\ \beta &= \frac{\alpha}{2\mu}.\end{aligned} \tag{3.66}$$

Putting $\{(1-\beta)^2+4\beta\rho\}^{\frac{1}{2}}=\gamma$ and calculating ΔL in the straightforward manner, and if $0<\beta<\frac{1}{2}$, then we get

$$\Delta L = -\frac{(1+\beta+\gamma)(3-\beta+\gamma)(1+\beta-\gamma)^2}{8(1-\beta+\gamma)} < 0. \tag{3.67}$$

Since the second term of L_C in (3.57) is the mean queue length, hence it is positive and $\xi < 0$ and $\rho < 1$,

$$(1 - 2\beta - \rho)\xi + 2\beta > 0. \tag{3.68}$$

Consequently, from (3.65) we find

$$L_C > L_I. \quad \square \tag{3.69}$$

3.4.2. Two-Stage Tandem Case

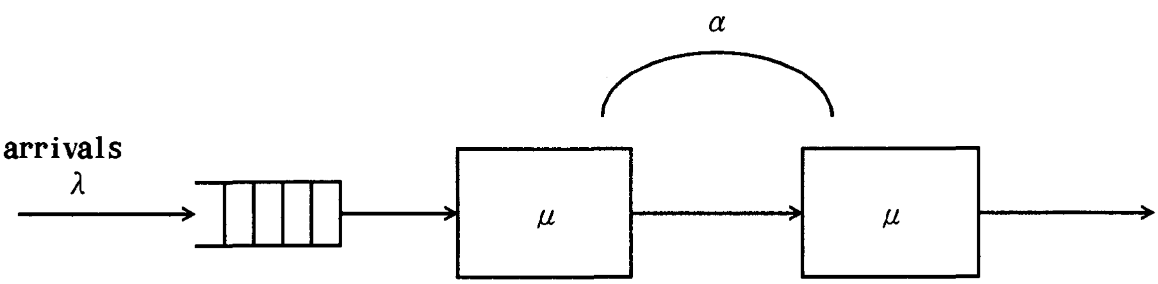

Fig. 3-4. Correlated two-stage tandem queueing system.

Fig. 3-4 shows the two-stage tandem queueing system with correlated service times described in Section 3.2. The service distribution of two stages is given by (3.1). We shall denote the mean number of customers in this system by $L_C(\mu, \mu, \alpha)$. From (3.22), it is given by

$$L_C(\mu, \mu, \alpha) = \frac{2(\mu+\alpha)^2(2\mu+\alpha)^2\lambda + \alpha(\mu+\alpha)(2\mu+\alpha)\lambda^2 - (4\mu^2+3\alpha\mu+\alpha^2)\lambda^3}{(\mu+\alpha)^2(2\mu+\alpha+\lambda)\{(\mu+\alpha)(2\mu+\alpha) - (3\mu+\alpha)\lambda\}}. \tag{3.70}$$

And the steady state condition is

$$\lambda < \frac{(\mu+\alpha)(2\mu+\alpha)}{3\mu+\alpha}. \tag{3.71}$$

On the other hand, we shall denote the mean number of customers in the ordinary two-stage tandem queueing system [86] by $L_I(\mu,\mu)$. It is given by

$$L_I(\mu,\mu) = \frac{4\rho(2-\rho^2)}{(2+\rho)(2-3\rho)}, \tag{3.72}$$

where

$$\rho = \frac{\lambda}{\mu}. \tag{3.73}$$

The steady state condition in the ordinary case [86] is $\rho < \frac{2}{3}$. Then, we do not have the same result as Theorem 3.3 as follows.

Theorem 3.4. *The mean number of customers $L_C(\mu-\alpha,\mu-\alpha,\alpha)$ in the two-stage tandem queueing system with correlated service times is smaller than $L_I(\mu,\mu)$ in the ordinary two-stage tandem queueing system if $0<\alpha<\mu$ and $\rho<\frac{2}{3}$.*

Proof. Replacing μ by $\mu-\alpha$ in (3.70), and from (3.72) we have

$$\begin{aligned} &L_I(\mu,\mu) - L_C(\mu-\alpha,\mu-\alpha,\alpha) \\ &\quad = \frac{\theta\rho^2[\{6\rho^3+3\rho^2+2(2-3\rho)\}(1-\theta)+(1+\rho)(\rho^2-2\rho+4)]}{(2+\rho)(2-3\rho)(2-\theta+\rho)\{(2-\theta)-\rho(3-2\theta)\}}, \end{aligned} \tag{3.74}$$

where

$$\theta = \frac{\alpha}{\mu}. \tag{3.75}$$

If $0<\theta<1$, then we have

$$\frac{2}{3} < \frac{2-\theta}{3-2\theta} < 1, \tag{3.76}$$

and when $\rho < \frac{2}{3}$ we find

$$(2-\theta) - \rho(3-2\theta) > 0. \tag{3.77}$$

Thus, the steady state condition of the correlated queueing system is satisfied.

Consequently, if $0 < \theta < 1$ and $\rho < \frac{2}{3}$ in (3.74), then

$$L_C(\mu-\alpha, \mu-\alpha, \mu) < L_I(\mu, \mu). \quad \square \tag{3.78}$$

3.5. Conclusion

We considered a tandem queueimg system with correlated service times whose distribution is the multivariate exponential distribution of Marshall and Olkin in this chapter. Both ordinary and commutative tandem queueing systems were discussed. For a two-stage ordinary tandem queueing system, we calculated the mean number of customers in the system by a generating function approach and showed analytically that it decreases according to the increasing of the correlation coefficient. On the other side, for a three-stage commutative tandem queueing system, we obtained by a matrix-geometric approach the throughput rates for three models which either an ordinary or commutative service operation and either of correlated or independent service times are combined and compared with one another. We established that for a small correlation coefficient, the throughput rate of the system is lower in the case of an ordinary service operation and correlated service times than in the case of a commutative service operation and independent service times, but for a large correlation coefficient, the reverse is true. We showed that the crossover between these two modes of behavior occurs at the ratio θ

of (approximately) 0.045, where θ equals the correlated service rate per two channels divided by the independent service rate per one channel.

Finally, we considered a correlated two-server parallel queueing system discussed in Section 2.2 and a correlated two-stage tandem queueing system discussed in Section 3.2. We showed that in the case of a two-server parallel system, the mean number of customers is greater when service times are correlated than when service times are independent, but in the case of a two-stage tandem system, the reverse is true, if the system can finish 2μ customers' service per unit time on the average.

Chapter 4.

INTERCHANGEABLE PARALLEL TANDEM QUEUEING SYSTEMS WITH CORRELATED SERVICE TIMES

4.1. Introduction

Nishida et al. [75] obtained the optimal allocation of service rates for a two-stage ordinary tandem queueing system with a single station, where optimality means that this allocation of service rates minimizes the rate of loss calls. Nishida [79] considered an interchangeable parallel two-stage tandem queueing system with no waiting room, where each stage consists of two stations in parallel. He called the discipline that a customer who finished the first-stage station can enter both of the second-stage stations as an interchangeable queueing system. He found the optimal allocation of service rates for the first and second stage stations in the sense of minimizing the rate of loss calls.

In this chapter, as an extension of the model introduced by Nishida [79], we consider an interchangeable parallel two-stage tandem queueing system in which the service times of each two stations in the first and second stage are assumed to follow the bivariate exponential distribution of Marshall and Olkin. For this system we get the throughput of the system using a matrix-geometric approach and compare the throughput of the system with those of an ordinary parallel two-stage tandem queueing system and an interchangeable parallel two-stage tandem queueing system whose service distribution is exponential.

A closely related model is that of Latouche and Neuts [54]. They considered a two-stage ordinary tandem queueing system with a feedbackloop, where each stage consists of a number of stations and one immediate finite waiting room exists between stages. They showed that their stationary probability vector has a matrix-geometric form.

An approach is closely related to Heavey et al. [35], Papadopoulos et al. [81,82] and Papadppoulos and O'Kelly [83]. They described the queueing system by a quasi-birth-death process and obtained the exact procedure of calculating the throughput. Similarly to the definition by the above four papers, the throughput treated in this chapter is defined as a critical input rate on the assumption that the first queue is never empty.

In Section 4.2, we describe our model fully, calculate the throughput rates four models and compare with the four throughput rates one another. In Section 4.3, we give a summary of Chapter 4.

4.2. Analysis of Model

4.2.1. Model

We consider a two-stage tandem queueing system with no intermediate buffers, in which each stage consists of two channels in parallel. Customers arrive at the queue, which is assumed to be infinite, according to a homogeneous Poisson process with rate $\lambda > 0$. After completion of the first stage service, each customer can enter both of the second stage channels. Namely, when the second stage channel which is located on as the same line as his first stage channel is busy, he can enter another channel in the second stage if it is free. We call this operation *interchangeable.* On the other side, after completion of the first stage service, each customer can enter only the second stage channel which is located on as the same line as his first stage channel if it is free. We call this operation *ordinary.* If a customer has already completed service of two stages, then he departs the system. But if he has not completed service of the second stage and his second stage channel is not free, he has to stay there, that is to say, this channel is blocked, and when his second stage channel has completed service, he can enter that channel. In the interchangeable case, a blocking occurs only when both of the second stage channels are busy, whereas in the ordinary case, it always occurs when the second stage channel which is located on as the same line as his first stage channel is busy.

The service times of each two channels in the first and second stages are not independent but depend upon each other. It is assumed that their service distribution is the bivariate exponential distribution of Marshall and Olkin. We call this model *correlated.* On the other side, we call a model in which their service distribution is an usual exponential distribution *independent.*

It is assumed that customers can transfer between channels instantaneously. Service to a customer at any channel, once initiated, is completed without interruptions. The queueing discipline is first-come first-served.

As shown in Fig. 4-1, the model I is interchangeable and correlated. We denote the throughput of this system by $\lambda^*_{I,C}$. Similarly, the model II, III and IV are interchangeable and independent, ordinary and correlated and ordinary and independent.

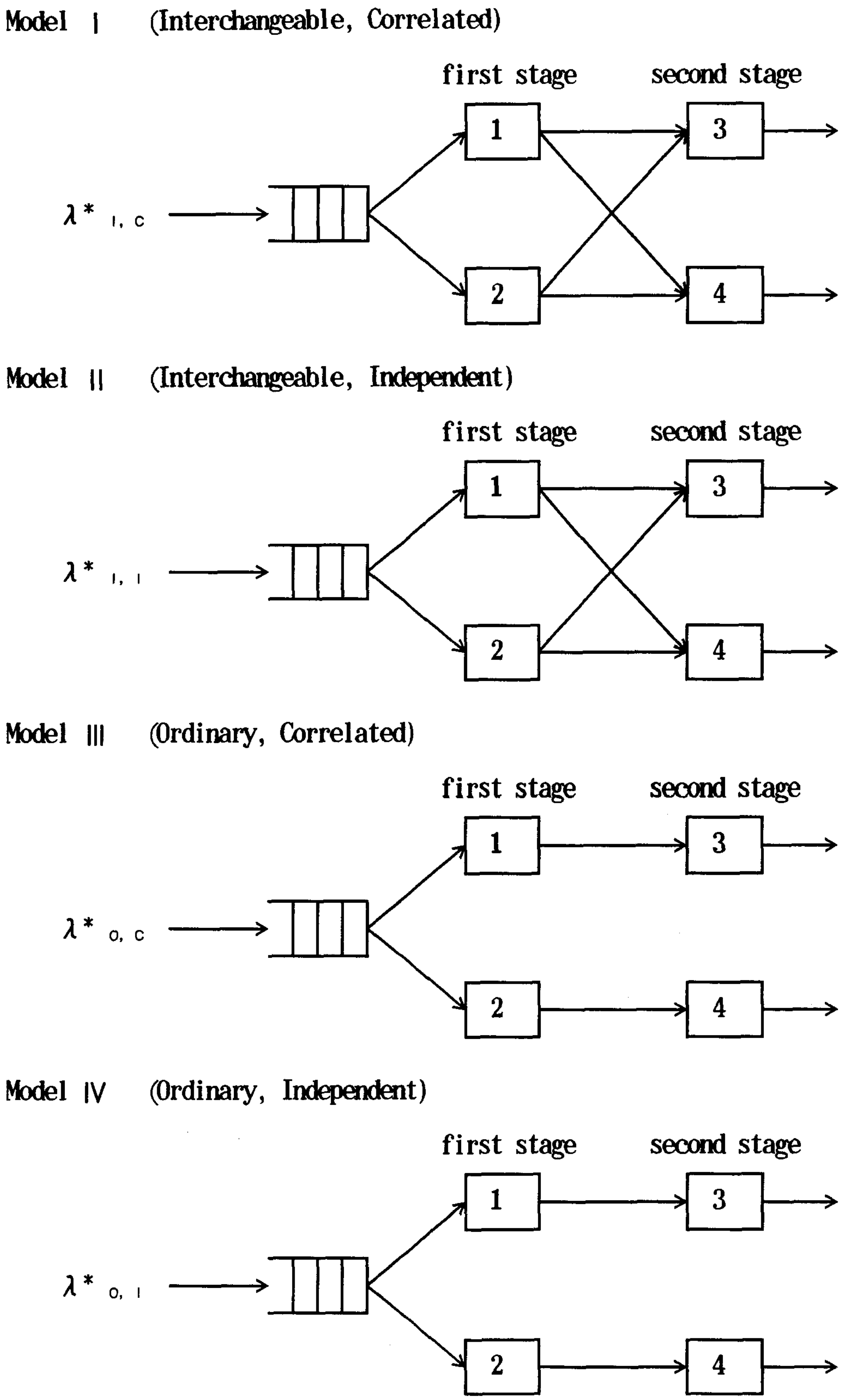

Fig. 4-1. Interchangeable parallel two-stage tandem queueing system.

We denote the throughput of the model II, III and IV by $\lambda^*_{I,I}$, $\lambda^*_{O,C}$ and $\lambda^*_{O,I}$, respectively.

4.2.2. Calculation of Throughput

We give numbers 1, 2, 3 and 4 to channels in the first and second stage as shown in Fig. 4-1. The joint service distribution of channel 1 and 2 and channel 3 and 4 are respectively given as follows:

$$\begin{aligned} \overline{F_1}(t_1, t_2) &= e^{-\mu_1 (t_1+t_2) - \alpha \max(t_1, t_2)}, \\ \overline{F_2}(t_3, t_4) &= e^{-\mu_2 (t_3+t_4) - \alpha \max(t_3, t_4)}. \end{aligned} \quad (4.1)$$

The model I under consideration can be studied as a continuous time Markov chain with state-space $\{(0, j) : 1 \le j \le m_0\} \cup \{(i, j) : i \ge 0, 1 \le j \le m\}$. The state $(0, j)$ denotes that the system is in the boundary state and m_0 denotes the number of the boundary states. In the state (i, j), i denotes the number of customers in the queue, whereas j denotes the state of the network consisting of four channels and m denotes the number of states in the network.

The states of the network are described by the vector:

$$(s_1, s_2, s_3, s_4) \quad (4.2)$$

where s_i $(i = 1, 2, 3, 4)$ can take any value from 0 to 2 with

$$s_i = \begin{cases} 0 & \text{-the } i\text{-th channel is idle,} \\ 1 & \text{-the } i\text{-th channel is in service,} \\ 2 & \text{-the } i\text{-th channel is blocked.} \end{cases} \quad (4.3)$$

Then the boundary states are

$$(0,0,0,0),(1,0,0,0),(0,1,0,0),(0,0,1,0),(0,0,0,1),$$
$$(0,0,1,1),(1,0,1,0),(1,0,0,1),(0,1,1,0),(0,1,0,1),$$
$$(1,0,1,1),(0,1,1,1),(0,2,1,1),(2,0,1,1). \tag{4.4}$$

The states of the network are

$$(1,1,0,0),(1,1,1,0),(1,1,0,1),(1,1,1,1),(2,1,1,1),$$
$$(1,2,1,1),(2,2,1,1). \tag{4.5}$$

By ordering the states as described above, the infinitesimal generator of the continuous time Markov chain has the following block partitioned structure:

$$\mathbf{Q} = \begin{bmatrix} \mathbf{A_{01}} & \mathbf{A_{04}} & & & & & \\ \mathbf{A_{02}} & \mathbf{A_1} & \mathbf{A_0} & & & & \\ \mathbf{A_{03}} & \mathbf{A_2} & \mathbf{A_1} & \mathbf{A_0} & & & \\ & \mathbf{A_3} & \mathbf{A_2} & \mathbf{A_1} & \mathbf{A_0} & & \\ & & \mathbf{A_3} & \mathbf{A_2} & \mathbf{A_1} & \mathbf{A_0} & \\ & & & \mathbf{A_3} & \mathbf{A_2} & \mathbf{A_1} & \\ & & & & & \vdots & \ddots \end{bmatrix} \tag{4.6}$$

where all the unmarked entries are zeros.

The dimensionality of $\mathbf{A_{01}}$ is 14×14, $\mathbf{A_{02}}$ and $\mathbf{A_{03}}$ are 7×14, $\mathbf{A_{04}}$ is 14×7, $\mathbf{A_0}$, $\mathbf{A_1}$, $\mathbf{A_2}$ and $\mathbf{A_3}$ are 7×7.

Let $\mathbf{p}$ equal the steady state probabilities of the network, assuming that the queue is never empty, which has elements $p(j)$ $(1 \leq j \leq 7)$. We can determine $\mathbf{p}$, by solving the system

$$\mathbf{pA} = \mathbf{0}, \qquad \mathbf{pe} = 1 \tag{4.7}$$

where the conservative matrix is given by

$$\mathbf{A} = \mathbf{A_0} + \mathbf{A_1} + \mathbf{A_2} + \mathbf{A_3} \tag{4.8}$$

and $\mathbf{e}$ is (7×1) column vector with all elements equal 1. The equilibrium condition is given (see Neuts [71]) by

$$\mathbf{pA_0e} < \mathbf{p}\,(\mathbf{A_2} + 2\mathbf{A_3})\,\mathbf{e}. \tag{4.9}$$

From this relationship the critical input rate, λ^*, to the system can be determined. In the steady state, this critical input rate is identical to the maximum throughput rate of the system.

The conservative matrix of the model I becomes

$$\mathbf{A} = \begin{bmatrix} -u_1 & 2\mu_1 & 0 & \alpha & 0 & 0 & 0 \\ u_2-\mu_2 & -v_1 & 0 & 2\mu_1 & 0 & \alpha & 0 \\ u_2-\mu_2 & 0 & -v_1 & 2\mu_1 & 0 & \alpha & 0 \\ \alpha & \mu_2 & \mu_2 & -v_1-\mu_2 & \mu_1 & \mu_1 & \alpha \\ 0 & \alpha & 0 & 2\mu_2 & -v_2 & 0 & u_1-\mu_1 \\ 0 & \alpha & 0 & 2\mu_2 & 0 & -v_2 & u_1-\mu_1 \\ 0 & 0 & 0 & \alpha & \mu_2 & \mu_2 & -u_2 \end{bmatrix}, \tag{4.10}$$

where

$$u_1 = 2\mu_1 + \alpha,$$

$$u_2 = 2\mu_2 + \alpha,$$

$$v_1 = 2\mu_1 + \mu_2 + 2\alpha,$$

$$v_2 = \mu_1 + 2\mu_2 + 2\alpha. \tag{4.11}$$

When $\mu_1 = \mu_2 = \mu$ and $\theta = \alpha/\mu$, the equilibrium condition (4.9) in this case is given by

$$\lambda < \mu\,(2+\theta)\,[\,1 - p(4)\,] + \mu\,\theta\,[\,p(1) + p(7)\,]. \tag{4.12}$$

By solving the system (4.7) and substituting $p(j)$ into (4.12), if $\mu = 1$, we obtain

$$\lambda < \frac{4(1+\theta)^2(6+\theta)}{16 + 21\theta + 3\theta^2} = \lambda^*_{I,C}. \tag{4.13}$$

When $\theta = 0$, that is, the service times of each two channels in the first and second stage are independent, the maximum throughput rate (4.13) yields

$$\lambda^*_{I,I} = \frac{3}{2}, \tag{4.14}$$

which gives the maximum throughput rate of the model II.

Next, we shall calculate the maximum throughput rate of the model III.

As additional boundary states, the states $(2,0,1,0)$ and $(0,2,0,1)$ are combined with the boundary states of the model I. Similarly, the states $(2,1,1,0)$ and $(1,2,0,1)$ are combined with the states of the network of the model I. The infinitesimal generator of the continuous time Markov chain has the same block partitioned structure as that of the model I. The dimensionality of $\mathbf{A_{01}}$ is 16×16, $\mathbf{A_{02}}$ and $\mathbf{A_{03}}$ are 9×16, $\mathbf{A_{04}}$ is 16×9, $\mathbf{A_0}$, $\mathbf{A_1}$, $\mathbf{A_2}$ and $\mathbf{A_3}$ are 9×9.

Let $\mathbf{p}$ equal the steady state probabilities of the network, assuming that the queue is never empty, which has elements $p(j)$ $(1 \le j \le 9)$. The relationship (4.7), (4.8) and (4.9) also hold for the model III.

The conservative matrix of the model III becomes

$$\mathbf{A} =$$
$$\begin{bmatrix} -u_1 & \mu_1 & \mu_1 & \alpha & 0 & 0 & 0 & 0 & 0 \\ u_2-\mu_2 & -v_1 & 0 & \mu_1 & \mu_1 & 0 & \alpha & 0 & 0 \\ u_2-\mu_2 & 0 & -v_1 & \mu_1 & 0 & \mu_1 & 0 & \alpha & 0 \\ \alpha & \mu_2 & \mu_2 & -v_1-\mu_2 & 0 & 0 & \mu_1 & \mu_1 & \alpha \\ 0 & u_2-\mu_2 & 0 & 0 & -v_{12} & 0 & u_1-\mu_1 & 0 & 0 \\ 0 & 0 & u_2-\mu_2 & 0 & 0 & -v_{12} & 0 & u_1-\mu_1 & 0 \\ 0 & \alpha & 0 & \mu_2 & \mu_2 & 0 & -v_2 & 0 & u_1-\mu_1 \\ 0 & 0 & \alpha & \mu_2 & 0 & \mu_2 & 0 & -v_2 & u_1-\mu_1 \\ 0 & 0 & 0 & \alpha & 0 & 0 & \mu_2 & \mu_2 & -u_2 \end{bmatrix} \tag{4.15}$$

where

$$v_{12} = \mu_1 + \mu_2 + 2\alpha. \tag{4.16}$$

When $\mu_1 = \mu_2 = \mu$ and $\theta = \alpha/\mu$, the equilibrium condition (4.9) in this case is given by

$$\lambda < \mu\,(1+\theta)\,[\,1 + p(1) + p(5) + p(6) + p(9) - p(4)\,]. \tag{4.17}$$

By solving the system (4.7) and substituting $p(j)$ into (4.17), if $\mu = 1$, we obtain

$$\lambda < \frac{4(1+\theta)}{3} = \lambda^*_{O,C}. \tag{4.18}$$

When $\theta = 0$, that is, the service times of each two channels in the first and second stage are independent, the maximum throughput rate (4.18) yields

$$\lambda^*_{O,I} = \frac{4}{3}, \tag{4.19}$$

which gives the maximum throughput rate of the model IV.

4.2.3. Comparison of Throughput

If

$$\theta = \theta_0 = 0.125, \tag{4.20}$$

then

$$\lambda^*_{I,I} = \lambda^*_{O,C}. \tag{4.21}$$

Consequently, since both $\lambda^*_{I,C}$ and $\lambda^*_{O,C}$ are increasing function with respect to $\theta > 0$, we obtain the following result.

Theorem 4.1. *If* $0 < \theta \leq \theta_0$, *then*

$$\lambda^*_{O,I} < \lambda^*_{O,C} \leq \lambda^*_{I,I} < \lambda^*_{I,C}. \tag{4.22}$$

And if $\theta > \theta_0$, then

$$\lambda^*_{O,I} < \lambda^*_{I,I} < \lambda^*_{O,C} < \lambda^*_{I,C}. \tag{4.23}$$

The following explanation is derived from the above theorem. When the correlation parameter is small, an interchangeable parallel two-stage tandem queueing model with independent service times improves the throughput of the system better than an ordinary parallel two-stage tandem queueing model with correlated service times, by utilizing an empty channel. When the correlation parameter is large, an ordinary parallel two-stage tandem queueing model with correlated service times improves it better than an interchangeable parallel two-stage tandem queueing model with independent service times, by the frequent occurrence of the event that two channels finish service at the same time.

4.3. Conclusion

We considered an interchangeable parallel two-stage tandem queueing system in which the service times of two channels in the first and second stage follow the bivariate exponential distribution of Marshall and Olkin in this chapter. The throughput rates were obtained by a matrix-geometric approach for four models which either of an ordinary or interchangeable service operation and either of correlated or independent service times are combined and compared with one another. It was a proper result that the throughput of a model with an interchangeable service operation and correlated service times is the greatest of the four and that of a model with a ordinary service operation and independent service times is the lowest of the four. More important result was that for a small correlation coefficient, the throughput of the system is lower in the case of an ordinary service operation and correlated service times than in the case of an interchangeable service operation and independent service times, but for a large correlation coefficient, the reverse is true. We showed that the crossover between these two modes of behavior occurs at the ratio θ of (approximately) 0.125, where θ equals the correlated service rate per two channels divided by the independent service rate per one channel.

Chapter 5.

TANDEM QUEUEING SYSTEMS WITH SEVERAL TYPES OF CUSTOMERS

5.1. Introduction

A common assumption made when analyzing tandem queueing systems is that the capacity of each station is infinite. That is, a station can accommodate any number of customers waiting to receive service. However, in real-life systems frequently the storage space in front of a station is finite. Due to this limitations imposed on the capacity, the flow of customers through one station may be momentarily stopped if a destination station has reached its capacity. This is known as blocking, and in view of this, a tandem queueing system with finite capacity is referred to as a tandem queue with blocking. A tandem queue with blocking have proved useful in modelling production systems, computer systems, and telecommunication systems (see Buzacott and Shanthikumar [8], and Papadopoulos et al. [84]).

Tandem queueing systems with finite capacity in which blocking and starvation are important is difficult to calculate performance measures explicitly because it has large state spaces (see Disney and König [19], and Perros [86]). Consequently, considerable attention has been focussed on the development of approximations whose method is commonly used by decomposition for evaluating performance measures. Several researchers, including Dijk and Lamond [17] and Shanthikumar and Jafari [93] have developed bounds for performance measures of tandem queueing systems with finite capacity. Another attention has been focussed on the establishment of convexity or concavity of performance measures (see Shaked and Shanthikumar [92]). There have been attempts to establish proof about similar problems for tandem queueing systems. Independently, Anantharam and Tsoucas [2] and Meester and Shanthikumar [59] demonstrated that the throughput of a tandem queueing system with finite capacity is increasing and concave with respect to the vector of capacity sizes.

On the other hand, in a conventional tandem queueing system, each customer requires several sequential types of service, each of which is performed by a different server who attends at each station. However, there are several practical examples of tandem queueing systems served by a single server: a labor and machine limited production system [69], a repairable system with a repair man [100], and operating systems in computer and telephone switching systems [39,40,41,64]. All the queues are served by a single server that moves among the stations according to a cyclic switching rule. That is , when the queueing system has N stages in series, a station S_i and a queue with infinite capacity, the server advances to the next station in cyclic order, $S_1 \rightarrow S_2 \rightarrow \cdots \rightarrow S_N \rightarrow S_1 \rightarrow$, and so on. The following switching rules [99] will be considered for each station.

(a) *Exhaustive service,* also called *a zero switching rule* : when the server visits a queue, its customers are served until that queue is empty.

(b) *K-limited service,* also called *a non-zero switching rule* : when the server visits a queue, it is served until either the queue becomes empty, or at most, a fixed number of customers, say K, are served, whichever occurs first.

(c) *Gated service* : when the server visits a queue, the customers found upon arrival at the queue are served.

Moreover, the server's walking time (sometimes known as switchover time or overhead time) is required, whenever the server moves from one station to the other.

Katayama [40] analyzed a two-stage tandem queueing system with walking time, which is served according to a gated service. Katayama and Kobayashi [43] and Nair [67] treated a two-stage tandem queueing system with zero switchover time and a K-limited service. Katayama [41] investigated a two-stage tandem queueing system with walking time and a K-limited service. Nair [66,68] and Taube-Netto [100] analyzed a two-stage tandem queueing system with zero switchover time and a exhaustive service. Murakami and Nakamura [64] studied a three-stage tandem queueing system with zero switchover time and a exhaustive service. Katayama [42] and König and Schmidt [48] studied a multi-stage tandem queueing system with zero switchover time, which is served according to several switching rules.

In Section 5.2, we investigate a two-stage ordinary tandem queueing system with two types of customers having independent Poisson arrivals and exponential service times, and obtain the mean queue length. An unlimited queue is allowed before the first stage but no queues are permitted between two stages. This second restriction results in the blocking of the first stage, whenever a customer who requires service in both stages having completed his service in the first stage cannot move into the second stage due to the presence of another customer there.

In Section 5.3, we consider a two-stage commutative tandem queueing system in which other assumption are the same as the model which will be discussed in Section 5.2. The blocking phenomenon occurs whenever a customer who requires service in both stages having completed his service in one stage but has to stay there because the other stage is still occupied.

In Section 5.4, we analyze a multi-stage tandem queueing system with zero switchover time, which is served according to a exhaustive service. Each arriving customer requires a certain number of sequential types of service, all of which are performed by a single server. We calculate the mean number of customers in the system and compare with the result in the case where each customer receives service in all stages.

In Section 5.5, we consider a two-stage tandem queueing system with no queues ahead of the first stage and finite intermediate buffer storage spaces. Each stage has a single station and the service times are independent and exponentially distributed with different parameters. For this system, we obtain the mean number of busy service stations, the rate of loss calls and the mean number of customers explicitly and establish tha concavity or the convexity of these performance measures. In Section 5.6, we have a summary of this chapter.

5.2. Two-Stage Tandem Queueing System with Two Types of Customers

5.2.1. Model

Fig. 5-1 shows a ordinary two-stage tandem queueing system. Two types of customers, namely type 1 and type 2 customers, arrive randomly and independently at stage 1. Interarrival times are exponentially distributed within classes with arrival rates λ_1 and λ_2. Each type 1 customer requires only one type of service and leaves the system without moving into stage 2 when he has completed his service in stage 1. Each type 2 customer requires two sequential types of service and leaves the system after completing his service in both stages. The service time distributions in the first and second stages are also exponential with service rates μ_1 and μ_2, respectively, and the service rate of type 1 customers in the first stage is the same as that of type 2 and a queue discipline is first-come first-served. An unlimited queue is allowed before stage 1 and no queues are permitted between stage 1 and stage 2. This second restriction results in the blocking of stage 1 whenever a type 2 customer having completed his service in stage 1 cannot move into stage 2 due to the presence of another type 2 customer there. If type 1 customers do not arrive at the system described above, it reduces to the ordinary two-stage tandem queueing system.

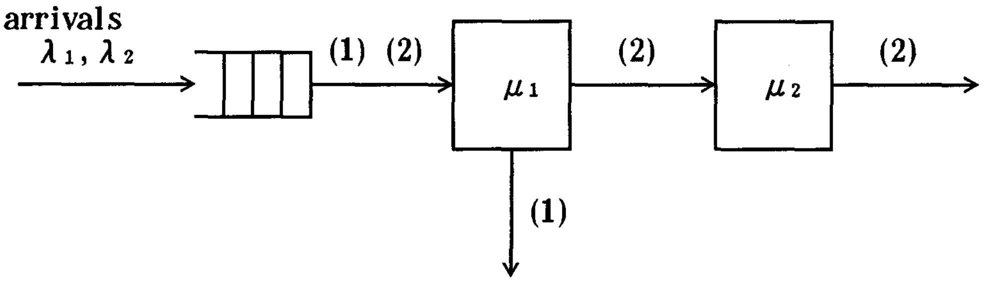

Fig. 5-1. Two-stage tandem queueing system with two types of customers.

5.2.2. Mean Queue Length

We define the following steady state probabilities:

$p_i^{(n)}(1,0)$ $(n \geq 0, i = 1,2)$: n customers in the queue and a type i customer in service in stage 1;

$p_i^{(n)}(1,1)$ $(n \geq 0, i = 1,2)$: n customers in the queue, a type i customer in service in stage 1 ,and a type 2 customer in service in stage 2;

$p^{(n)}(b,1)$ $(n \geq 0)$: n customers in the queue, a type 2 customer in stage 1 with service completed, and a type 2 customer in service in stage 2;

p_0 : no customers in the system ;

p_1 : no customers in the queue or in stage 1 and a type 2 customer in service in stage 2.

The steady state equations may be constructed in the usual manner. They are

$$(\lambda_1 + \lambda_2)p_0 = \mu_1 p_1^{(0)}(1,0) + \mu_2 p_1,$$

$$(\lambda_1 + \lambda_2 + \mu_2)p_1 = \mu_1 p_1^{(0)}(1,1) + \mu_1 p_2^{(0)}(1,0) + \mu_2 p^{(0)}(b,1),$$

$$(\lambda_1 + \lambda_2 + \mu_1)p_1^{(0)}(1,0) = \lambda_1 p_0 + \mu_2 p_1^{(0)}(1,1) + \mu_1 \beta_1 p_1^{(1)}(1,0),$$

$$(\lambda_1 + \lambda_2 + \mu_1)p_2^{(0)}(1,0) = \lambda_2 p_0 + \mu_2 p_2^{(0)}(1,1) + \mu_1 \beta_2 p_1^{(1)}(1,0),$$

$$(\lambda_1 + \lambda_2 + \mu_2)p^{(0)}(b,1) = \mu_1 p_2^{(0)}(1,1),$$

$$(\lambda_1 + \lambda_2 + \mu_1 + \mu_2)p_1^{(0)}(1,1) = \lambda_1 p_1 + \mu_1 \beta_1 p_1^{(1)}(1,1) + \mu_2 \beta_1 p^{(1)}(b,1) + \mu_1 \beta_1 p_2^{(1)}(1,0),$$

$$(\lambda_1 + \lambda_2 + \mu_1 + \mu_2)p_2^{(0)}(1,1) = \lambda_2 p_1 + \mu_1 \beta_2 p_1^{(1)}(1,1) + \mu_2 \beta_2 p^{(1)}(b,1) + \mu_1 \beta_2 p_2^{(1)}(1,0),$$

$$\begin{aligned}(\lambda_1 + \lambda_2 + \mu_1)p_1^{(n)}(1,0) &= (\lambda_1 + \lambda_2)p_1^{(n-1)}(1,0) + \mu_2 p_1^{(n)}(1,1) \\ &\quad + \mu_1 \beta_1 p_1^{(n+1)}(1,0), \qquad n \geq 1,\end{aligned}$$

$$\begin{aligned}(\lambda_1 + \lambda_2 + \mu_1)p_2^{(n)}(1,0) &= (\lambda_1 + \lambda_2)p_2^{(n-1)}(1,0) + \mu_2 p_2^{(n)}(1,1) \\ &\quad + \mu_1 \beta_2 p_1^{(n+1)}(1,0), \qquad n \geq 1,\end{aligned}$$

$$(\lambda_1+\lambda_2+\mu_2)p^{(n)}(b,1)=(\lambda_1+\lambda_2)p^{(n-1)}(b,1)+\mu_1 p_2^{(n)}(1,1), \qquad n\geq 1,$$

$$(\lambda_1+\lambda_2+\mu_1+\mu_2)p_1^{(n)}(1,1)=(\lambda_1+\lambda_2)p_1^{(n-1)}(1,1)+\mu_1\beta_1 p_1^{(n+1)}(1,1)$$
$$+\mu_2\beta_1 p^{(n+1)}(b,1)+\mu_1\beta_1 p_2^{(n+1)}(1,0), \qquad n\geq 1,$$

$$(\lambda_1+\lambda_2+\mu_1+\mu_2)p_2^{(n)}(1,1)=(\lambda_1+\lambda_2)p_2^{(n-1)}(1,1)+\mu_1\beta_2 p_1^{(n+1)}(1,1)$$
$$+\mu_2\beta_2 p^{(n+1)}(b,1)+\mu_1\beta_2 p_2^{(n+1)}(1,0), \qquad n\geq 1. \tag{5.1}$$

We introduce the following generating functions:

$$F_i(z)=\sum_{n=0}^{\infty} p_i^{(n)}(1,0)z^n, \qquad i=1,2,$$
$$F_3(z)=\sum_{n=0}^{\infty} p^{(n)}(b,1)z^n,$$
$$F_j(z)=\sum_{n=0}^{\infty} p_{j-3}^{(n)}(1,1)z^n, \qquad j=4,5. \tag{5.2}$$

Multiplying the equations (5.1) by appropriate powers of z and adding them up, we have

$$\{\rho z^2-\nu z+\beta_1\}F_1(z)+\gamma z F_4(z)=\beta_1\rho(1-z)p_0-\beta_1\gamma p_1,$$
$$\beta_2 F_1(z)+z(\rho z-\nu)F_2(z)+\gamma z F_5(z)=\beta_2\rho(1-z)p_0-\beta_2\gamma p_1,$$
$$\{\rho z-(\rho+\gamma)\}F_3(z)+F_5(z)=0,$$
$$\beta_1 F_2(z)+\beta_1\gamma F_3(z)+\{\rho z^2-(\gamma+\nu)z+\beta_1\}F_4(z)=\beta_1(\gamma+\rho-\rho z)p_1,$$
$$\beta_2 F_2(z)+\beta_2\gamma F_3(z)+\beta_2 F_4(z)+z\{\rho z-(\gamma+\nu)\}F_5(z)=\beta_2(\gamma+\rho-\rho z)p_1, \tag{5.3}$$

where

$$\lambda=\lambda_1+\lambda_2, \quad \beta_1=\frac{\lambda_1}{\lambda}, \quad \beta_2=\frac{\lambda_2}{\lambda}, \quad \gamma=\frac{\mu_2}{\mu_1}, \quad \rho=\frac{\lambda}{\mu_1}, \quad \nu=\rho+1. \tag{5.4}$$

From (5.3), we obtain

$$F_i(z) = \frac{H_i(z)}{\Delta(z)}, \qquad 1 \le i \le 5, \tag{5.5}$$

where

$$\Delta(z) =$$

$$\begin{vmatrix} \rho z^2 - \nu z + \beta_1 & 0 & 0 & \gamma z & 0 \\ \beta_2 & z(\rho z - \nu) & 0 & 0 & \gamma z \\ 0 & 0 & \rho z - (\rho + \gamma) & 0 & 1 \\ 0 & \beta_1 & \beta_1 \gamma & \rho z^2 - (\gamma + \nu) z + \beta_1 & 0 \\ 0 & \beta_2 & \beta_2 \gamma & \beta_2 & z\{\rho z - (\gamma + \nu)\} \end{vmatrix}$$

$$= \mu_1^5 z^2 (z-1)(\rho z - \nu)\{\rho z - (\gamma+\nu)\}[\rho^3 z^4 - 2\rho^2(\gamma+\nu) z^3$$

$$+ \rho\{\rho^2 + 2(\beta_1 + \gamma + 1)\rho + \gamma^2 + 3\gamma + 1\} z^2 - \{2\beta_1 \rho^2 + (2\beta_1 + 2\gamma + \beta_1 \gamma)\rho$$

$$+ \gamma(\gamma+1)\} z + \beta_1(\beta_1 \rho + \gamma)], \tag{5.6}$$

and $H_i(z)$ is the determinant obtained from $\Delta(z)$ by replacing the i-th column by a column vector

$$\begin{bmatrix} \beta_1 \rho (1-z) p_0 - \beta_1 \gamma p_1 \\ \beta_2 \rho (1-z) p_0 - \beta_2 \gamma p_1 \\ 0 \\ \beta_1 (\gamma + \rho - \rho z) p_1 \\ \beta_2 (\gamma + \rho - \rho z) p_1 \end{bmatrix}. \tag{5.7}$$

From (5.6), it is easily verified that

$$\left[\frac{H_1(z)}{1-z}\right]_{z=1} > 0. \tag{5.8}$$

Hence, since

$$0 < F_1(1) \le 1, \tag{5.9}$$

we obtain

$$\left[\frac{\Delta(z)}{1-z}\right]_{z=1} > 0, \tag{5.10}$$

namely

$$\rho < \frac{\beta_2\gamma + \gamma^2}{\beta_2^2 + \beta_2\gamma + \gamma^2}. \tag{5.11}$$

Note that for $\beta_2 = 1$

$$\rho < \frac{\gamma + \gamma^2}{1 + \gamma + \gamma^2}. \tag{5.12}$$

In this case the above system reduces to that of the ordinary two-stage tandem case with different service rates.

Then, allowable utilization factor ρ of each queueing system is given as follows;

$$\text{Ordinary system:} \qquad 0 < \rho < \frac{1}{1 + 1/(\gamma + \gamma^2)}, \tag{5.13}$$

$$\text{This system:} \qquad 0 < \rho < \frac{1}{1 + \beta_2^2/(\beta_2\gamma + \gamma^2)}. \tag{5.14}$$

Therefore, it follows that compared with the ordinary case, the maximum utilization of this system is increased when $0 < \beta_2 < 1$.

Now the equation

$$\begin{aligned} &\rho^3 z^4 - 2\rho^2(\gamma+\nu)z^3 + \rho\{\rho^2 + 2(\beta_1+\gamma+1)\rho + \gamma^2 + 3\gamma + 1\}z^2 \\ &- \{2\beta_1\rho^2 + (2\beta_1 + 2\gamma + \beta_1\gamma)\rho + \gamma(\gamma+1)\}z + \beta_1(\beta_1\rho + \gamma) = 0 \end{aligned} \tag{5.15}$$

has one and only one real root in $0 < z < 1$ under the condition (5.11). Hence it turns out that $\Delta(z) = 0$ has one and only one real root in $0 < |z| < 1$ since the equation

$$(\rho z - \nu)\{\rho z - (\gamma + \nu)\} = 0 \tag{5.16}$$

has no root whose absolute value is less than unity.

Let us denote this root by η. Since the generating functions $F_i(z) = H_i(z)/\Delta(z) (1 \le i \le 5)$ are regular in the domain $|z| \le 1$, the root of the numerator and the denominator of the right-hand side must coincide with each other. Therefore, it follows that

$$H_i(z) = 0, \qquad 1 \le i \le 5. \tag{5.17}$$

Using the relation

$$\begin{aligned}&\rho^3\eta^4 - 2\rho^2(\gamma+\nu)\eta^3 + \rho\{\rho^2 + 2(\beta_1+\gamma+1)\rho + \gamma^2 + 3\gamma + 1\}\eta^2 \\ &- \{2\beta_1\rho^2 + (2\beta_1 + 2\gamma + \beta_1\gamma)\rho + \gamma(\gamma+1)\}\eta + \beta_1(\beta_1\rho + \gamma) = 0,\end{aligned} \tag{5.18}$$

we notice that when any one of equations (5.17) holds, another four equations follow from it. This shows that a relation between p_0 and p_1 is given as follows;

$$\beta_2\rho p_0 - \{\rho^2\eta^2 - \rho(\gamma+\nu)\eta + \beta_1\rho + \gamma\}p_1 = 0. \tag{5.19}$$

Taking account of the total probability, it holds that

$$p_0 + p_1 + \sum_{i=1}^{5} F_i(1) = 1. \tag{5.20}$$

From the equations (5.3), we have

$$\begin{aligned}\beta_2F_1'(1) - \gamma F_4'(1) &= (\rho-1)F_1(1) + \gamma F_4(1) + \beta_1\rho p_0,\\ -\beta_2F_1'(1) + F_2'(1) - \gamma F_5'(1) &= (\rho-1)F_2(1) + \gamma F_5(1) + \beta_2\rho p_0,\\ \gamma F_3'(1) - F_5'(1) &= \rho F_3(1),\\ -\beta_1F_2'(1) - \beta_1\gamma F_3'(1) + (\beta_2+\gamma)F_4'(1) &= (\rho-\gamma-1)F_4(1) + \beta_1\rho p_1,\\ -\beta_2F_2'(1) - \beta_2\gamma F_3'(1) - \beta_2F_4'(1) + (\gamma+1)F_5'(1) &= (\rho-\gamma-1)F_5(1) + \beta_2\rho p_1.\end{aligned} \tag{5.21}$$

Addition of these equations and substitution of the normalization condition (5.20) yield

$$F_1(1) + F_2(1) + F_4(1) + F_5(1) = \rho. \tag{5.22}$$

Similarly from the equations (5.3), we have

$$\begin{aligned}
\beta_2 F_1''(1) - \gamma F_4''(1) =& 2\rho F_1(1) + 2(\rho - 1)F_1'(1) \\
& + 2\gamma F_4'(1), \\
-\beta_2 F_1''(1) + F_2''(1) - \gamma F_5''(1) =& 2\rho F_2(1) + 2(\rho - 1)F_2'(1) \\
& + 2\gamma F_5'(1), \\
\gamma F_3''(1) - F_5''(1) =& 2\rho F_3'(1), \\
-\beta_1 F_2''(1) - \beta_1\gamma F_3''(1) + (\beta_2 + \gamma)F_4''(1) =& 2\rho F_4(1) + 2(\rho - \gamma - 1)F_4'(1), \\
-\beta_2 F_2''(1) - \beta_2\gamma F_3''(1) - \beta_2 F_4''(1) + (\gamma + 1)F_5''(1) =& 2\rho F_5(1) + 2(\rho - \gamma - 1)F_5'(1).
\end{aligned} \tag{5.23}$$

Addition of these equations and substitution of the relation (5.22) yield

$$(\rho - 1)F_1'(1) + (\rho - 1)F_2'(1) + \rho F_3'(1) + (\rho - 1)F_4'(1) + (\rho - 1)F_5'(1) = -\rho^2. \tag{5.24}$$

Moreover, from the first four equations of (5.3), it is evident that $F_i(1)(1 \le i \le 5)$ satisfy the following relations;

$$\begin{aligned}
\beta_2 F_1(1) - \gamma F_4(1) &= \beta_1 \gamma p_1, \\
-\beta_2 F_1(1) + F_2(1) - \gamma F_5(1) &= \beta_2 \gamma p_1, \\
\gamma F_3(1) - F_5(1) &= 0, \\
-\beta_1 F_2(1) - \beta_1 \gamma F_3(1) + (\beta_2 + \gamma) F_4(1) &= -\beta_1 \gamma p_1.
\end{aligned} \tag{5.25}$$

Solving the above equations and (5.22) for $F_i(1)(1 \leq i \leq 5)$ in the usual way, we have

$$F_1(1) = \frac{\beta_1\gamma(\rho + p_1)}{\beta_2 + \gamma}, \qquad F_2(1) = \frac{\beta_2 F_1(1)}{\beta_1}, \qquad F_3(1) = \frac{\beta_2(\beta_2\rho - \gamma p_1)}{\gamma(\beta_2 + \gamma)},$$

$$F_4(1) = \frac{\beta_1\gamma F_3(1)}{\beta_2}, \qquad F_5(1) = \gamma F_3(1). \tag{5.26}$$

Substituting the above results into the normalization condition (5.20), we get the following relation between p_0 and p_1;

$$\gamma(\beta_2 + \gamma)p_0 + \gamma^2 p_1 = \beta_2\gamma + \gamma^2 - \rho(\beta_2^2 + \beta_2\gamma + \gamma^2). \tag{5.27}$$

Thus, by (5.19) and (5.27), we find that

$$p_0 = \frac{\{\beta_2\gamma + \gamma^2 - \rho(\beta_2^2 + \beta_2\gamma + \gamma^2)\}\{\rho^2\eta^2 - \rho(\gamma + \nu)\eta + \beta_1\rho + \gamma\}}{\gamma(\beta_2 + \gamma)\{\rho^2\eta^2 - \rho(\gamma + \nu)\eta + \beta_1\rho + \gamma\} + \beta_2\rho\gamma^2},$$

$$p_1 = \frac{\beta_2\rho p_0}{\rho^2\eta^2 - \rho(\gamma + \nu)\eta + \beta_1\rho + \gamma}. \tag{5.28}$$

Note that for $\beta_2 = \gamma = 1$

$$\eta = 0, \tag{5.29}$$

and therefore

$$p_0 = \frac{2 - 3\rho}{2 + \rho}, \tag{5.30}$$

which coincides with that of the ordinary case [86]. Because, in this case the equation (5.15) yields

$$z\, h(z) = 0, \tag{5.31}$$

where

$$h(z) = \rho^3 z^3 - 2\rho^2(\rho + 2)z^2 + \rho(\rho^2 + 4\rho + 5)z - 2(\rho + 1). \tag{5.32}$$

Now we shall derive the expected number of customers in the queue. In order to determine this, we must obtain $F_i'(1)(1 \le i \le 5)$ explicitly. The first four equations of (5.21) are four such equations and the equation (5.24) is remaining. These equations can be solved for $F_i'(1)(1 \le i \le 5)$ by elementary but a little troublesome calculations and the mean queue length L_q is given by

$$
\begin{aligned}
L_q &= \sum_{i=1}^{5} F_i'(1) \\
&= \frac{1}{(\gamma+1)\{\beta_2\gamma+\gamma^2-\rho(\beta_2^2+\beta_2\gamma+\gamma^2)\}} \\
&\cdot [(1-\rho)(\beta_2\gamma+2\beta_2+\gamma)F_1(1)+(1-\rho)(\beta_2+\gamma)F_2(1) \\
&+\rho(\beta_2^2+\beta_2\gamma+\gamma^2+\beta_2^2\gamma)F_3(1)+\{\beta_2+\gamma-\rho(\beta_2\gamma+\beta_2+\gamma)\}F_4(1) \\
&-\gamma(\beta_2+\gamma)F_5(1)+\rho^2(\gamma+1)(\beta_2^2+\beta_2\gamma+\gamma^2)+\rho(\beta_2^2\gamma+\beta_2^2-\beta_2\gamma-2\beta_2-\gamma)p_0 \\
&-\beta_1\rho(\beta_2\gamma+\beta_2+\gamma)p_1]. \qquad (5.33)
\end{aligned}
$$

The result (5.33) holds, of course, for $\beta_2 = \gamma = 1$. Let $\beta_2 = \gamma = 1$ in (5.26), then

$$
F_1(1) = F_4(1) = 0,
$$

$$
F_2(1) = \frac{\rho(2-\rho)}{2+\rho}, \qquad F_3(1) = F_5(1) = \frac{2\rho^2}{2+\rho}. \qquad (5.34)
$$

Thus, from (5.30) the result (5.33) is

$$
L_q = \frac{4\rho^2(1+2\rho)}{(2+\rho)(2-3\rho)}, \qquad (5.35)
$$

which is identical with the mean queue length in the ordinary case.

5.3. Two-Stage Commutative Tandem Queueing System with Two Types of Customers

5.3.1. Model

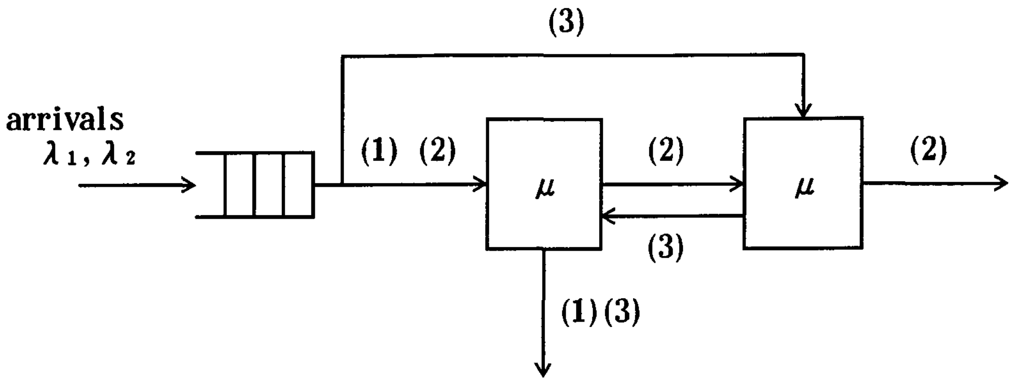

Fig. 5-2. Commutative tandem queueing system with two types of customers.

Fig. 5-2 shows a two-stage commutative tandem queueing system, where type 1 and type 2 customers arrive according to a Poisson stream with parameter λ_1 and λ_2, respectively. Each type 1 customer requires only one service, can receive service in either stage and leaves the system without moving into the other stage when he has completed service in one stage. In a word, type 1 customers intend to behave as like in the $M/M/2(\infty)$ queueing system. Each type 2 customer requires two types of service and leaves the system after completing service in both stages. Each type 2 customer on arrival enters stage 1 if both stages are free, and enter stage 2 if stage 1 is busy. In a word, type 2 customers intend to behave as like in the usual commutative tandem queueing system as mentioned above. Each service time of both stages is exponentially distributed with parameter μ. Infinite queue is possible ahead of stage 1, but no queues are permitted between both stages. The second restriction causes the phenomenon of blocking which means the situation that takes place when a type 2 customer has completed service in one stage but has to stay there because the other stage is still occupied. Each type 2 customer transfer between stages instantaneously if other stage is vacant. The queueing discipline is first-come first -served. If type 1 or type 2 customers do not arrive at the system, it reduces to the two-stage commutative tandem queueing system with infinite waiting rooms or the $M/M/2(\infty)$ queueing system, respectively.

5.3.2. Mean Queue Length

The notation $p(s_1, s_2, s_3)$ will be used to denote the steady state probability, where s_1 is the length of queue, s_2 is the state of stage 1 and s_3 is the state of stage 2. We define the following labels to denote the state of each stage.

0 : no customers,

1 : a type 1 customer in service,

2 : a type 2 customer in the first service,

3 : a type 2 customer in the second service,

b : a type 2 customer is blocked.

Viewing the nature of this system, the steady state equations can be easily constructed in the usual manner. But, for simplicity, we shall set as follows:

$$p_0 = p(0,0,0),$$

$$p_1 = p(0,1,0) + p(0,0,1),$$

$$p_2 = p(0,2,0) + p(0,0,2),$$

$$p_3 = p(0,3,0) + p(0,0,3),$$

$$p(n,1) = p(n,1,1), \qquad n \geq 0,$$

$$p(n,2) = p(n,2,2), \qquad n \geq 0,$$

$$p(n,3) = p(n,3,3), \qquad n \geq 0,$$

$$p(n,4) = p(n,1,2) + p(n,2,1), \qquad n \geq 0,$$

$$p(n,5) = p(n,2,3) + p(n,3,2), \qquad n \geq 0,$$

$$p(n,6) = p(n,1,3) + p(n,3,1), \qquad n \geq 0,$$

$$p(n,7) = p(n,1,b) + p(n,b,1), \qquad n \geq 0,$$

$$p(n,8) = p(n,2,b) + p(n,b,2), \qquad n \geq 0,$$

$$p(n,9) = p(n,3,b) + p(n,b,3), \qquad n \geq 0. \tag{5.36}$$

Then the steady state equations are written in a simple manner as follows:

$$\lambda p_0 + \mu p_1 + \mu p_3 = 0,$$

$$-(\lambda + \mu)p_1 + \lambda_1 p_0 + 2\mu p(0,1) + \mu p(0,6) = 0,$$

$$-(\lambda + \mu)p_2 + \lambda_2 p_0 + \mu p(0,4) + \mu p(0,5) = 0,$$

$$-(\lambda + \mu)p_3 + \mu p(0,6) + \mu p_2 + \mu p(0,7) + \mu p(0,9) + 2\mu p(0,3) = 0,$$

$$-(\lambda + 2\mu)p(0,1) + \lambda_1 p_1 + \mu\beta_1 p(1,6) + 3\mu\beta_1 p(1,1) = 0,$$

$$-(\lambda + 2\mu)p(0,2) + \lambda_2 p_2 + \mu\beta_2 p(1,4) + \mu\beta_2 p(1,5) = 0,$$

$$-(\lambda + 2\mu)p(0,3) + \mu p(0,8) = 0,$$

$$-(\lambda + 2\mu)p(0,4) + \lambda_1 p_2 + \lambda_2 p_1 + \mu\beta_1 p(1,4) + \mu\beta_1 p(1,5) + 2\mu\beta_2 p(1,1)$$
$$+\mu\beta_2 p(1,6) = 0,$$

$$-(\lambda+2\mu)p(0,5)+\lambda_2 p_3+\mu\beta_2 p(1,6)+2\mu\beta_2 p(1,3)+\mu\beta_2 p(1,7)$$
$$+\mu\beta_2 p(1,9)=0,$$
$$-(\lambda+2\mu)p(0,6)+\lambda_1 p_3+\mu\beta_1 p(1,6)+2\mu\beta_1 p(1,3)+\mu\beta_1 p(1,9)$$
$$+\mu\beta_1 p(1,7)=0,$$
$$-(\lambda+\mu)p(0,7)+\mu p(0,4)=0,$$
$$-(\lambda+\mu)p(0,8)+2\mu p(0,2)=0,$$
$$-(\lambda+\mu)p(0,9)+\mu p(0,5)=0,$$
$$-(\lambda+2\mu)p(n,1)+\lambda p(n-1,1)+\mu\beta_1 p(n+1,6)+2\mu\beta_1 p(n+1,1)=0, \quad n\geq 1,$$
$$-(\lambda+2\mu)p(n,2)+\lambda p(n-1,2)+\mu\beta_2 p(n+1,4)+\mu\beta_2 p(n+1,5)=0, \quad n\geq 1,$$
$$-(\lambda+2\mu)p(n,3)+\lambda p(n-1,3)+\mu p(n,8)=0, \quad n\geq 1,$$
$$-(\lambda+2\mu)p(n,4)+\lambda p(n-1,4)+\mu\beta_1 p(n+1,4)+\mu\beta_1 p(n+1,5)$$
$$+2\mu\beta_2 p(n+1,1)+\mu\beta_2 p(n+1,6)=0, \quad n\geq 1,$$
$$-(\lambda+2\mu)p(n,5)+\lambda p(n-1,5)+\mu\beta_2 p(n+1,6)+2\mu\beta_2 p(n+1,3)$$
$$+\mu\beta_2 p(n+1,7)+\mu\beta_2 p(n+1,9)=0, \quad n\geq 1,$$
$$-(\lambda+2\mu)p(n,6)+\lambda p(n-1,6)+\mu\beta_1 p(n+1,6)+2\mu\beta_1 p(n+1,3)$$
$$+\mu\beta_1 p(n+1,9)+\mu\beta_1 p(n+1,7)=0, \quad n\geq 1,$$
$$-(\lambda+\mu)p(n,7)+\lambda p(n-1,7)+\mu p(n,4)=0, \quad n\geq 1,$$
$$-(\lambda+\mu)p(n,8)+\lambda p(n-1,8)+2\mu p(n,2)=0, \quad n\geq 1,$$
$$-(\lambda+\mu)p(n,9)+\lambda p(n-1,9)+\mu p(n,5)=0, \quad n\geq 1, \tag{5.37}$$

where

$$\lambda = \lambda_1 + \lambda_2, \qquad \beta_1 = \frac{\lambda_1}{\lambda}, \qquad \beta_2 = \frac{\lambda_2}{\lambda}. \tag{5.38}$$

We introduce the following generating functions:

$$F_i(z) = \sum_{n=0}^{\infty} p(n,i) z^n, \qquad 1 \le i \le 9. \tag{5.39}$$

Multiplying the equations (5.37) by appropriate powers of z and adding them up, we have

$$\begin{aligned}
(fz + 2\beta_1)F_1(z) + \beta_1 F_6(z) &= -\beta_1^2 \rho p_0 - \beta_1 g p_1,\\
fzF_2(z) + \beta_2 F_4(z) + \beta_2 F_5(z) &= -\beta_2^2 \rho p_0 - \beta_2 g p_2,\\
fF_3(z) + F_8(z) &= 0,\\
2\beta_2 F_1(z) + (fz + \beta_1)F_4(z) + \beta_1 F_5(z) + \beta_2 F_6(z) &= -2\beta_1\beta_2 \rho p_0 - \beta_2 g p_1 - \beta_1 g p_2,\\
2\beta_2 F_3(z) + fzF_5(z) + \beta_2 F_6(z) + \beta_2 F_7(z) + \beta_2 F_9(z) &= -\beta_2 p_2 - \beta_2 g p_3,\\
2\beta_1 F_3(z) + (fz + \beta_1)F_6(z) + \beta_1 F_7(z) + \beta_1 F_9(z) &= -\beta_1 p_2 - \beta_1 g p_3,\\
F_4(z) + gF_7(z) &= 0,\\
2F_2(z) + gF_8(z) &= 0,\\
F_5(z) + gF_9(z) &= 0,
\end{aligned} \tag{5.40}$$

where

$$f = \rho z - (\rho + 2), \qquad g = \rho z - (\rho + 1), \qquad \rho = \frac{\lambda}{\mu}. \tag{5.41}$$

From (5.40), we obtain

$$F_i(z) = \frac{H_i(z)}{\Delta(z)}, \qquad 1 \le i \le 9, \tag{5.42}$$

where

$$\Delta(z) = \begin{vmatrix} fz+2\beta_1 & 0 & 0 & 0 & 0 & \beta_1 & 0 & 0 & 0 \\ 0 & fz & 0 & \beta_2 & \beta_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & f & 0 & 0 & 0 & 0 & 1 & 0 \\ 2\beta_2 & 0 & 0 & fz+\beta_1 & \beta_1 & \beta_2 & 0 & 0 & 0 \\ 0 & 0 & 2\beta_2 & 0 & fz & \beta_2 & \beta_2 & 0 & \beta_2 \\ 0 & 0 & 2\beta_1 & 0 & 0 & fz+\beta_1 & \beta_1 & 0 & \beta_1 \\ 0 & 0 & 0 & 1 & 0 & 0 & g & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & g & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & g \end{vmatrix}.$$

$$= z^2(z-1)f^2g^2(fz+\beta_1)[\rho^4 z^5 - \rho^3(3\rho+7)z^4$$

$$+ \rho^2\{3\rho^2 + (3\beta_1+14)\rho + 18\}z^3 - \rho\{\rho^3 + (6\beta_1+7)\rho^2$$

$$+ (14\beta_1+19)\rho + 20\}z^2 + \{3\beta_1\rho^3 + (2\beta_1^2 + 14\beta_1 + 1)\rho^2$$

$$+ (20\beta_1+4)\rho + 8\}z - (2\beta_1^2\rho^2 + 6\beta_1^2\rho + 8\beta_1 - 4)], \tag{5.43}$$

and $H_i(z)(1 \leq i \leq 9)$ is the determinant obtained from $\Delta(z)$ by replacing the i-th column by a column vector:

$$\begin{bmatrix} -\beta_1^2\rho p_0 - \beta_1 g p_1 \\ -\beta_2^2\rho p_0 - \beta_2 g p_2 \\ 0 \\ -2\beta_1\beta_2\rho p_0 - \beta_2 g p_1 - \beta_1 g p_2 \\ -\beta_2 p_2 - \beta_2 g p_3 \\ -\beta_1 p_2 - \beta_1 g p_3 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{5.44}$$

From (5.42), if the system is in steady state,

$$0 < F_i(1) < 1, \qquad 1 \leq i \leq 9. \tag{5.45}$$

Therefore, the steady state condition should be

$$\rho < \frac{2(3-2\beta_1)}{(2-\beta_1)(4-3\beta_1)}. \tag{5.46}$$

Substituting $\beta_1 = 0$ in this inequality, we get

$$\rho < \frac{3}{4}, \tag{5.47}$$

which is the steady state condition for the usual commutative tandem queueing system [76].

On the other hand, let $\beta_1 = 1$ in (5.46), then

$$\rho < 2 \quad \text{i.e.,} \quad \frac{\lambda}{2\mu} < 1, \tag{5.48}$$

which coincides with that of the $M/M/2(\infty)$ queueing system.

Now the equation

$$fz + \beta_1 = 0 \quad \text{i.e.,} \quad \rho z^2 - (\rho+2)z + \beta_1 = 0 \tag{5.49}$$

has one and only one real root in $0 < z < 1$, namely

$$\kappa = \frac{\rho + 2 - (\rho^2 + 4\beta_2\rho + 4)^{\frac{1}{2}}}{2\rho}. \tag{5.50}$$

Also, the equation

$$\begin{aligned}
&\rho^4 z^5 - \rho^3(3\rho+7)z^4 + \rho^2\{3\rho^2 + (3\beta_1 + 14)\rho + 18\}z^3 - \rho\{\rho^3 + (6\beta_1 + 7)\rho^2 \\
&+ (14\beta_1 + 19)\rho + 20\}z^2 + \{3\beta_1\rho^3 + (2\beta_1^2 + 14\beta_1 + 1)\rho^2 + (20\beta_1 + 4)\rho + 8\}z \\
&- (2\beta_1^2\rho^2 + 6\beta_1^2\rho + 8\beta_1 - 4) = 0
\end{aligned} \tag{5.51}$$

has more than one real root in $|z| < 1$ under the condition (5.46). Let η be one of such roots. Hence it turns out that the equation $\Delta(z) = 0$ has more than two real roots in $|z| < 1$ since the equation

$$f^2 g^2 = 0 \quad \text{i.e.,} \quad \{\rho z - (\rho+1)\}^2\{\rho z - (\rho+2)\}^2 = 0 \tag{5.52}$$

has no root whose absolute value is less than unity. Since the generating functions (5.42) are regular in the domain $|z| \leq 1$, the root of numerator and the denominator of the right-hand side must coincide each other. Therefore,

$$H_i(\kappa) = 0, \qquad 1 \leq i \leq 9, \tag{5.53}$$

and

$$H_i(\eta) = 0, \qquad 1 \leq i \leq 9. \tag{5.54}$$

Using the relation

$$\rho\kappa^2 - (\rho+2)\kappa + \beta_1 = 0, \tag{5.55}$$

and

$$\begin{aligned}
&\rho^4\eta^5 - \rho^3(3\rho+7)\eta^4 + \rho^2\{3\rho^2 + (3\beta_1+14)\rho + 18\}\eta^3 - \rho\{\rho^3 + (6\beta_1+7)\rho^2 \\
&+ (14\beta_1+19)\rho + 20\}\eta^2 + \{3\beta_1\rho^3 + (2\beta_1^2 + 14\beta_1 + 1)\rho^2 + (20\beta_1+4)\rho + 8\}\eta \\
&- (2\beta_1^2\rho^2 + 6\beta_1^2\rho + 8\beta_1 - 4) = 0,
\end{aligned} \tag{5.56}$$

we notice that when any one equation holds, another eight equations follow from it. This shows that two relations are given as follows:

$$\beta_2 p_1 - \beta_1 p_2 = 0, \tag{5.57}$$

and

$$\begin{aligned}
&\beta_2\rho\{\rho^2\eta^2 - \rho(2\rho+3)\eta + \rho^2 + (\beta_1+2)\rho + 2\}p_0 + \{\rho\eta - (\rho+1)\}\{\rho^2\eta^2 \\
&- \rho(\rho+3)\eta + 2(\beta_1\rho+1)\}p_2 = 0.
\end{aligned} \tag{5.58}$$

Taking account of the total probability, we have

$$\sum_{i=0}^{3} p_i + \sum_{j=1}^{9} F_j(1) = 1. \tag{5.59}$$

Now, we shall derive the mean queue length.

Differentiating both sides of (5.40) with regard to z and then substituting $z = 1$, we have

$$\begin{aligned}
-2\beta_2 F_1'(1) + \beta_1 F_6'(1) &= (2-\rho)F_1(1) - \beta_1 \rho p_1, \\
-2F_2'(1) + \beta_2 F_4'(1) + \beta_2 F_5'(1) &= (2-\rho)F_2(1) - \beta_2 \rho p_2, \\
-2F_3'(1) + F_8'(1) &= -\rho F_3(1), \\
2\beta_2 F_1'(1) + (\beta_1 - 2)F_4'(1) + \beta_1 F_5'(1) + \beta_2 F_6'(1) &= (2-\rho)F_4(1) - \beta_2 \rho p_1 - \beta_1 \rho p_2, \\
2\beta_2 F_3'(1) - 2F_5'(1) + \beta_2 F_6'(1) + \beta_2 F_7'(1) + \beta_2 F_9'(1) &= (2-\rho)F_5(1) - \beta_2 \rho p_3, \\
2\beta_1 F_3'(1) + (\beta_1 - 2)F_6'(1) + \beta_1 F_7'(1) + \beta_1 F_9'(1) &= (2-\rho)F_6(1) - \beta_1 \rho p_3, \\
F_4'(1) - F_7'(1) &= -\rho F_7(1), \\
2F_2'(1) - F_8'(1) &= -\rho F_8(1), \\
F_5'(1) - F_9'(1) &= -\rho F_9(1).
\end{aligned} \tag{5.60}$$

Addition of these equations and substitution of the normalization condition (5.59) yield

$$F_1(1) + F_2(1) + F_4(1) + F_5(1) + F_6(1) = \frac{1}{2}\rho(1 - p_0). \tag{5.61}$$

Similarly from the equations (5.40), we have

$$\begin{aligned}
-2\beta_2 F_1''(1) + \beta_1 F_6''(1) &= -2\rho F_1(1) - 2(\rho - 2)F_1'(1), \\
-2F_2''(1) + F_4''(1) + F_5''(1) &= -2\rho F_2(1) - 2(\rho - 2)F_2'(1),
\end{aligned}$$

$$-2F_3''(1) - F_8''(1) = -2\rho F_3'(1),$$

$$2\beta_2 F_1''(1) + (\beta_1 - 2)F_4''(1) + \beta_1 F_5''(1) + \beta_2 F_6''(1) = -2\rho F_4(1) - 2(\rho - 2)F_4'(1),$$

$$2\beta_2 F_3''(1) - 2F_5''(1) + \beta_2 F_6''(1) + \beta_2 F_7''(1) + \beta_2 F_9''(1) = -2\rho F_5(1) - 2(\rho - 2)F_5'(1),$$

$$2\beta_1 F_3''(1) + (\beta_1 - 2)F_6''(1) + \beta_1 F_7''(1) + \beta_1 F_9''(1) = -2\rho F_6(1) - 2(\rho - 2)F_6'(1),$$

$$F_4''(1) - F_7''(1) = -2\rho F_7'(1),$$

$$2F_2''(1) - F_8''(1) = -2\rho F_8'(1),$$

$$F_5''(1) - F_9''(1) = -2\rho F_9'(1). \tag{5.62}$$

Addition of these equations and substitution of the relation (5.61) yield

$$(\rho - 2)\{F_1'(1) + F_2'(1) + F_4'(1) + F_5'(1) + F_6'(1)\}$$
$$+ \rho\{F_3'(1) + F_7'(1) + F_8'(1) + F_9'(1)\} = -\frac{1}{2}\rho^2(1 - p_0). \tag{5.63}$$

Moreover, from the first four and the last four equations of (5.40), it is evident that equations $F_i(1)$ $(1 \leq i \leq 9)$ satisfy the following relations:

$$-2\beta_2 F_1(1) + \beta_1 F_6(1) = -\beta_1^2 \rho p_0 + \beta_1 p_1,$$

$$-2F_2(1) + \beta_2 F_4(1) + \beta_2 F_5(1) = -\beta_2^2 \rho p_0 + \beta_2 p_2,$$

$$-2F_3(1) + F_8(1) = 0,$$

$$2\beta_2 F_1(1) + (\beta_1 - 2)F_4(1) + \beta_1 F_5(1) + \beta_2 F_6(1) = -2\beta_1\beta_2 \rho p_0 + \beta_2 p_1 + \beta_1 p_2,$$

$$2\beta_1 F_3(1) + (\beta_1 - 2)F_6(1) + \beta_1 F_7(1) + \beta_1 F_9(1) = -\beta_1 p_2 + \beta_1 p_3,$$

$$F_4(1) - F_7(1) = 0,$$

$$2F_2(1) - F_8(1) = 0,$$

$$F_5(1) - F_9(1) = 0. \tag{5.64}$$

Solving the above equations and (5.61) for $F_i(1)$ $(1 \leq i \leq 9)$ in the usual way, we obtain $F_1(1), F_2(1)$ and $F_6(1)$ as follows:

$$F_1(1) = \frac{\beta_1^2\rho + \beta_1^2\rho p_0 - 2\beta_1 p_1}{2(3-2\beta_1)},$$

$$F_2(1) = \frac{\beta_2^2 F_1(1)}{\beta_1^2},$$

$$F_6(1) = \frac{\beta_1(1-\beta_1)\rho - \beta_1(2-\beta_1)\rho p_0 + p_1}{3-2\beta_1}. \tag{5.65}$$

From (5.59), (5.61) and (5.65), we get the following relation between p_0 and p_2:

$$\{(2-\beta_1^2)\rho + 2(3-2\beta_1)\}p_0 + 2\beta_2 p_2 = 2(3-2\beta_1) - (2-\beta_1)(4-3\beta_1)\rho. \tag{5.66}$$

Thus, by (5.58) and (5.66), we find

$$p_0 = \{2(3-2\beta_1) - (2-\beta_1)(4-3\beta_1)\rho\}\{\rho\eta - (\rho+1)\}\{\rho^2\eta^2 - \rho(\rho+3)\eta + 2(\beta_1\rho+1)\}/D, \tag{5.67}$$

where

$$\begin{aligned} D =& \{(2-\beta_1^2)\rho + 2(3-2\beta_1)\}\{\rho\eta - (\rho+1)\}\{\rho^2\eta^2 - \rho(\rho+3)\eta + 2(\beta_1\rho+1)\} \\ & - 2\beta_2^2\rho\{\rho^2\eta^2 - \rho(2\rho+3)\eta + \rho^2 + (\beta_1+2)\rho + 2\}. \end{aligned} \tag{5.68}$$

So letting $\beta_1 = 0$ and $\beta_2 = 1$, we have

$$p_0 = \{-(4\rho^4 - 3\rho^3)\eta^3 + (8\rho^4 + 10\rho^3 - 12\rho^2)\eta^2 - (4\rho^4 + 13\rho^3 + 8\rho^2 - 15\rho)\eta + 8\rho^2 + 2\rho - 6\}/D_0, \tag{5.69}$$

where

$$D_0 = (\rho^4 + 3\rho^3)\eta^3 - (2\rho^4 + 11\rho^3 + 12\rho^2)\eta^2 + (\rho^4 + 9\rho^3 + 20\rho^2 + 15\rho)\eta - (\rho^3 + 4\rho^2 + 10\rho + 6), \tag{5.70}$$

which agrees with the empty probability of the commutative tandem queue [76].

Now we shall derive the expected number of customers in the queue. In order to determine this, we must obtain $F_i'(1)$ $(1 \leq i \leq 9)$ explicitly. The first four and the last four equations of (5.60) are such eight equations and the equation (5.63) is remaining. These equations can be solved for $F_i'(1)$ $(1 \leq i \leq 9)$ by elementary but a little troublesome calculation and the mean queue length L_q is given by

$$\begin{aligned} L_q =&[(2-\beta_1)(4-3\beta_1)\rho^3 + (6\beta_1^3 - 21\beta_1^2 + 30\beta_1 - 20)\rho^3 p_0 + 4(1-\beta_1)(5-3\beta_1)\rho^2 p_2 \\ &+ 2(5-3\beta_1)(2-\rho)\rho F_1(1) + 2\{(9\beta_1^2 - 38\beta_1 + 37)\rho - 2(4-3\beta_1)\}\rho F_2(1) \\ &+ 2(5-3\beta_1)(2-\rho)\rho F_6(1)]/[2\{2(3-2\beta_1) - (2-\beta_1)(4-3\beta_1)\rho\}\rho]. \end{aligned} \tag{5.71}$$

Since type 1 customers don't arrive at the system in the case $\beta_1 = 0$, i.e., $\beta_2 = 1$, it should be noted that

$$F_1(1) = F_6(1) = 0. \tag{5.72}$$

Hence, letting $\beta_1 = 0$ and $\beta_2 = 1$ in (5.65) and (5.71), we have

$$L_q = \{(3\rho^2 - 6\rho + 16)p_0 - (39\rho^2 - 38\rho + 16)\}/\{4(4\rho - 3)\}, \tag{5.73}$$

which coincides with the mean queue length of the commutative tandem queue [76].

Similarly, since type 2 customers don't arrive at the system in the case $\beta_1 = 1$,

$$p_2 = F_2(1) = F_6(1) = 0. \tag{5.74}$$

Hence, let $\beta_1 = 1$ in (5.66), then

$$p_0 = \frac{2-\rho}{2+\rho} = \frac{2\mu - \lambda}{2\mu + \lambda}, \tag{5.75}$$

which coincides with the probability that the system is empty in the $M/M/2(\infty)$.

Moreover, letting $\beta_1 = 1$ in (5.65) and (5.71), we obtain

$$L_q = \frac{\rho^3}{(2+\rho)(2-\rho)}, \tag{5.76}$$

which coincides with the mean queue length of the $M/M/2(\infty)$.

5.4. n Queues in Tandem Attended by A Single Server

5.4.1. Model

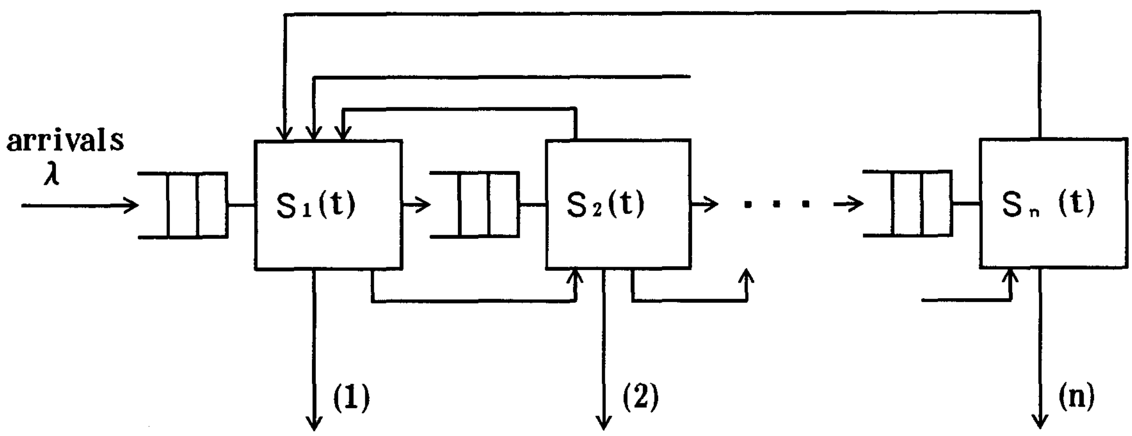

Fig. 5-3. Multi-stage tandem queueing system attended by a single server.

Fig. 5-3 shows an ordered sequence of n single service stations and n types of customers arrive according to a Poisson process of density λ. Independently of the others, an arriving customer is of type l $(1 \le l \le n)$ with probability β_l where $\sum_{l=1}^{n} \beta_l = 1$. Each customer of type l requires l sequential types of service. For example, each customer of type 1 requires only service in stage 1 and each customer of type n requires service in all stages. All of service are performed by a single server, who switches instantaneously from one stage of the system to another whenever the number of customers in the present stage is zero. The service time processes in each stage are assumed to constitute independent renewal processes with arbitrary distribution $S_k(t)$ $(1 \leq k \leq n)$, with finite moments. These processes are also assumed to be stochastically independent of the arrival and switching time processes. The queue capacities in front of service stations are assumed to be infinite. A customer arriving at the system waits for service in stage 1 unless the server is ready to serve him in that stage, in which case service immediately begins. When a customer of type 1 completes service in stage 1, he leaves the system. On the other hand, when any other type of customer completes service in stage 1, he immediately goes to stage 2 and joins queue 2. However, if the server moves with him to stage 2 and no other customer is already waiting there, then his service in stage 2 commences immediately. When the server attends stage 1, he takes two following switching way: if stage 1 is empty and queue 2 accumulates, then he switches to stage 2 and if the system is empty, then he remains in stage 1. Next, when the server attends stage 2, a customer of type 2 leaves the system and any other type of customers goes to stage 3 after completing service in stage 2. In this stage the server takes three following switching way: if stage 2 is empty and queue 3 accumulates, then he switches to stage 3 and if both stage 2 and stage 3 are empty and queue 1 accumulates, then he switches to stage 1 and if the system is empty, then he remains in stage 2. Similarly, when the server attends somewhere through stage 3 to stage $n-1$, the above fashion is also followed. Finally, when the server attends stage n, each customer leaves the system after the completion of his service. In this stage the server takes two following way: if stage n is empty and queue 1 accumulates, then he switches to stage 1 and if the system is empty, then he remains in stage n. When the server switches from stage n to stage 1, the above is repeated.

5.4.2. Mean Number of Customers

The following symbols will be used as follows;

S_k $(1 \leq k \leq n)$: a random variable representing service time in stage k,

$S_k(t)$ $(1 \leq k \leq n)$: the distribution of S_k,

$E[S_k]$ $(1 \leq k \leq n)$: the expected value of S_k,

$\tilde{S}_k(s)$ $(1 \leq k \leq n)$: the Laplace-Stieltjes transform of $S_k(t)$,

$\pi_l(m; i_1, i_2, \cdots, i_{n-1}, i_n)$ $(1 \leq k, l, m \leq n)$: the steady state probability that stage k has i_k customers just after a customer of type l has just completed service in stage m and leaves stage m,

$\beta_l^{(k)}$ $(1 \leq k \leq n, k \leq l \leq n)$: the probability that a customer of type l demands service in stage k,

$Q_{l,m}^{(k)}$ $(1 \leq k, m \leq n, m \leq l \leq n)$: a random variable representing the number of customers in stage k just after a customer of type l departs from stage m.

We also define

$$p_i^{(k)} = \int_0^\infty (i!)^{-1}(\lambda t)^i \exp(-\lambda t) dS_k(t), \qquad 1 \leq k \leq n, \tag{5.77}$$

as the probability of having i arrivals during a service time of a customer in stage k.

The corresponding generating function is

$$P_k(x) = \sum_{i=0}^{\infty} p_i^{(k)} x^i = \tilde{S}_k[\lambda(1-x)], \qquad 1 \leq k \leq n, \tag{5.78}$$

which converges for $|x| \le 1$. Note that the first and second derivatives of $P_k(x)$ at $x = 1$ are $\rho_k = \lambda E[S_k]$ and $\lambda^2 E[S_k^2]$, respectively.

Assuming that a steady state exists, we have the following balance equations;

$$\pi_1(1; i_1, 0, \cdots, 0) = p_{i_1}^{(1)} \beta_1^{(1)} \sum_{m=1}^{n} \pi_m(m; 0, \cdots, 0) + \sum_{a=1}^{i_1+1} p_{i_1-a+1}^{(1)} \beta_1^{(1)} \sum_{m=1}^{n} \pi_m(m; a, 0, \cdots, 0), \qquad i_1 \ge 0,$$

$$\pi_1(1; i_1, i_2, 0, \cdots, 0) = \sum_{a=1}^{i_1+1} p_{i_1-a+1}^{(1)} \beta_1^{(1)} \sum_{m=1}^{n} \pi_m(1; a, i_2, 0, \cdots, 0), \qquad i_1 \ge 0, i_2 \ge 1,$$

$$\pi_l(1; i_1, 1, 0, \cdots, 0) = p_{i_1}^{(1)} \beta_l^{(1)} \sum_{m=1}^{n} \pi_m(m; 0, \cdots, 0) + \sum_{a=1}^{i_1+1} p_{i_1-a+1}^{(1)} \beta_l^{(1)} \sum_{m=1}^{n} \pi_m(m; a, 0, \cdots, 0), \qquad 2 \le l \le n, i_1 \ge 0,$$

$$\pi_l(1; i_1, i_2, 0, \cdots, 0) = \sum_{a=1}^{i_1+1} p_{i_1-a+1}^{(1)} \beta_l^{(1)} \sum_{m=1}^{n} \pi_m(1; a, i_2 - 1, 0, \cdots, 0), \qquad 2 \le l \le n, i_1 \ge 0, i_2 \ge 2,$$

$$\pi_2(2; i_1, i_2, 0, \cdots, 0) = p_{i_1}^{(2)} \beta_2^{(2)} \sum_{m=1}^{n} \pi_m(1; 0, i_2 + 1, 0, \cdots, 0) + \sum_{a=0}^{i_1} p_{i_1-a}^{(2)} \beta_2^{(2)} \pi_2(2; a, i_2 + 1, 0, \cdots, 0), \qquad i_1, i_2 \ge 0,$$

$$\pi_2(2; i_1, i_2, i_3, 0, \cdots, 0) = \sum_{a=0}^{i_1} p_{i_1-a}^{(2)} \beta_2^{(2)} \sum_{m=2}^{n} \pi_m(2; a, i_2 + 1, i_3, 0, \cdots, 0), \qquad i_1, i_2 \ge 0, i_3 \ge 1,$$

$$\pi_l(2;i_1,i_2,1,0,\cdots,0)=p_{i_1}^{(2)}\beta_l^{(2)}\sum_{m=1}^{n}\pi_m(1;0,i_2+1,0,\cdots,0)$$

$$+\sum_{a=0}^{i_1}p_{i_1-a}^{(2)}\beta_l^{(2)}\pi_2(2;a,i_2+1,0,\cdots,0),$$

$$3\le l\le n,i_1,i_2\ge 0,$$

$$\pi_l(2;i_1,i_2,i_3,0,\cdots,0)=\sum_{a=0}^{i_1}p_{i_1-a}^{(2)}\beta_l^{(2)}\sum_{m=2}^{n}\pi_m(2;a,i_2+1,i_3-1,0,\cdots,0),$$

$$3\le l\le n,i_1,i_2\ge 0,i_3\ge 2,$$

$$\pi_k(k;i_1,0,\cdots,0,i_k,0,\cdots,0)$$

$$=\sum_{a=0}^{i_1}p_{i_1-a}^{(k)}\beta_k^{(k)}\sum_{m=k-1}^{n}\pi_m(k-1;a,0,\cdots,0,i_k+1,0,\cdots,0)$$

$$+\sum_{a=0}^{i_1}p_{i_1-a}^{(k)}\beta_k^{(k)}\pi_k(k;a,0,\cdots,0,i_k+1,0,\cdots,0),$$

$$3\le k\le n-1,i_1,i_k\ge 0,$$

$$\pi_k(k;i_1,0,\cdots,0,i_k,i_{k+1},0,\cdots,0)$$

$$=\sum_{a=0}^{i_1}p_{i_1-a}^{(k)}\beta_k^{(k)}\sum_{m=k}^{n}\pi_m(k;a,0,\cdots,0,i_k+1,i_{k+1},0,\cdots,0),$$

$$3\le k\le n-1,i_1,i_k\ge 0,i_{k+1}\ge 1,$$

$$\pi_l(k;i_1,0,\cdots,0,i_k,1,0,\cdots,0)$$

$$=\sum_{a=0}^{i_1}p_{i_1-a}^{(k)}\beta_l^{(k)}\sum_{m=k-1}^{n}\pi_m(k-1;a,0,\cdots,0,i_k+1,0,\cdots,0)$$

$$+\sum_{a=0}^{i_1}p_{i_1-a}^{(k)}\beta_l^{(k)}\pi_k(k;a,0,\cdots,0,i_k+1,0,\cdots,0),$$

$$3\le k\le n-1,k+1\le l\le n,i_1,i_k\ge 0,$$

$$\pi_l(k;i_1,0,\cdots,0,i_k,i_{k+1},0,\cdots,0)$$
$$= \sum_{a=0}^{i_1} p_{i_1-a}^{(k)} \beta_l^{(k)} \sum_{m=k}^{n} \pi_m(k;a,0,\cdots,0,i_k+1,i_{k+1}-1,0,\cdots,0),$$
$$3 \le k \le n-1, k+1 \le l \le n, i_1, i_k \ge 0, i_{k+1} \ge 2,$$

$$\pi_n(n;i_1,0,\cdots,0,i_n)$$
$$= \sum_{a=0}^{i_1} p_{i_1-a}^{(n)} \{\pi_{n-1}(n-1;a,0,\cdots,0,i_n+1)$$
$$+ \pi_n(n-1;a,0,\cdots,0,i_n+1) + \pi_n(n;a,0,\cdots,0,i_n+1)\}. \tag{5.79}$$

Also the normalization condition yields

$$\sum_{l=1}^{n} \sum_{m=1}^{n} \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} \cdots \sum_{i_n=0}^{\infty} \pi_l(m;i_1,i_2,\cdots,i_n) = 1. \tag{5.80}$$

We introduce the generating functions:

$$U_l(x,y) = \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} x^{i_1} y^{i_2} \pi_l(1;i_1,i_2,0,\cdots,0), \qquad 1 \le l \le n, |x| \le 1, |y| \le 1,$$

$$V_l^{(k)}(x,y,z) = \sum_{i_1=0}^{\infty} \sum_{i_k=0}^{\infty} \sum_{i_{k+1}=0}^{\infty} x^{i_1} y^{i_k} z^{i_{k+1}} \pi_l(k;i_1,0,\cdots,0,i_k,i_{k+1},0,\cdots,0),$$
$$2 \le k \le n-1, k \le l \le n, |x| \le 1, |y| \le 1, |z| \le 1,$$

$$W(x,z) = \sum_{i_1=0}^{\infty} \sum_{i_n=0}^{\infty} x^{i_1} z^{i_n} \pi_n(n;i_1,0,\cdots,0,i_n), \qquad |x| \le 1, |z| \le 1. \tag{5.81}$$

Then, the equations (5.79) give

$$U_l(x,y) = [\beta_l^{(1)} P_1(x) y^{1-\delta_{1,l}} \{\sum_{m=2}^{n-1} V_m^{(m)}(x,0,0) + W(x,0) + (x-1)E - \sum_{m=1}^{n} U_m(0,y) + U_1(0,0)\}]/[x - \{\sum_{j=1}^{n} \beta_j^{(1)} y^{1-\delta_{1,j}}\} P_1(x)],$$

$$1 \le l \le n,$$

$$V_l^{(2)}(x,y,z) = [\beta_l^{(2)} P_2(x) z^{1-\delta_{2,l}} \{\sum_{m=1}^{n} U_m(0,y) - U_1(0,0) - \sum_{m=2}^{n} V_m^{(2)}(x,0,z)\}]/[y - \{\sum_{j=2}^{n} \beta_j^{(2)} z^{1-\delta_{2,j}}\} P_2(x)],$$

$$2 \le l \le n,$$

$$V_l^{(k)}(x,y,z) = [\beta_l^{(k)} P_k(x) z^{1-\delta_{k,l}} \{\sum_{m=k-1}^{n} V_m^{(k-1)}(x,0,y) - V_{k-1}^{(k-1)}(x,0,0) - \sum_{m=k}^{n} V_m^{(k)}(x,0,z)\}]/[y - \{\sum_{j=k}^{n} \beta_j^{(k)} z^{1-\delta_{k,j}}\} P_k(x)],$$

$$3 \le k \le n-1, k \le l \le n,$$

$$W(x,z) = [P_n(x)\{V_{n-1}^{(n-1)}(x,0,z) + V_n^{(n-1)}(x,0,z) - V_{n-1}^{(n-1)}(x,0,0) - W(x,0)\}]/\{z - P_n(x)\}, \tag{5.82}$$

where

$$E = U_1(0,0) + \sum_{m=2}^{n-1} V_m^{(m)}(0,0,0) + W(0,0), \tag{5.83}$$

and $\delta_{i,j}$ is the Kronecker delta.

The condition (5.80) is

$$\sum_{l=1}^{n} U_l(1,1) + \sum_{k=2}^{n-1} \sum_{l=k}^{n} V_l^{(k)}(1,1,1) + W(1,1) = 1. \tag{5.84}$$

In order to characterize completely the distribution $\{\pi_l(m; i_1, i_2, \cdots, i_n)\}$ by the transforms $U_l(x,y), V_l^{(k)}(x,y,z)$ and $W(x,z)$, we need to determine the unknowns on the numerator of the right-hand side in each transform, but though we may not do that, we can attain our objectives by deriving some relation from its regularity in the domain.

Since each customer of type l $(1 \le l \le n)$ leaves the system after the completion of his service in stage l, we have the following relation about $\beta_l^{(k)}$:

$$\beta_l^{(k)} = \beta_l \Big/ \sum_{j=k}^{n} \beta_j, \qquad 1 \le k \le n, k \le l \le n,$$

$$\beta_l^{(1)} = \beta_l, \qquad 1 \le l \le n,$$

$$\beta_n^{(n)} = 1, \qquad \sum_{l=k}^{n} \beta_l^{(k)} = 1. \tag{5.85}$$

$U_l(1,1)$ $(1 \le l \le n)$ is obtained from $U_l(x,y)$ in (5.82) by applying L'Hopital's rule as follows:

$$U_l(1,1) = \lim_{y \to 1} U_l(1,y)$$

$$= \left[\beta_l \sum_{m=1}^{n} \{\frac{d}{dy} U_m(0,y)\}_{y=1}\right] \Big/ \sum_{j=2}^{n} \beta_j. \tag{5.86}$$

Similarly, applying L'Hopital's rule to $V_l^{(k)}(x,y,z)$ $(2 \le k \le n-1, k \le l \le n)$ and $W(x,z)$ and considering (5.85), we get

$$U_l(1,1) = V_l^{(2)}(1,1,1), \qquad 2 \le l \le n,$$

$$V_l^{(k-1)}(1,1,1) = V_l^{(k)}(1,1,1), \qquad 3 \le k \le n-1, k \le l \le n,$$

$$V_n^{(n-1)}(1,1,1) = W(1,1). \tag{5.87}$$

Then, the condition (5.84) is rewritten as

$$\sum_{l=1}^{n} l U_l(1,1) = 1. \tag{5.88}$$

Therefore, $U_l(1,1), V_l^{(k)}(1,1,1)$ and $W(1,1)$ is determined from (5.86), (5.87) and (5.88):

$$U_l(1,1) = \beta_l \Big/ \sum_{j=1}^{n} j\beta_j, \qquad 1 \leq l \leq n,$$

$$V_l^{(k)}(1,1,1) = \beta_l \Big/ \sum_{j=1}^{n} j\beta_j, \qquad 2 \leq k \leq n-1, k \leq l \leq n-1,$$

$$V_n^{(k)}(1,1,1) = W(1,1) = \beta_n \Big/ \sum_{j=1}^{n} j\beta_j, \quad 2 \leq k \leq n-1. \tag{5.89}$$

The probability that the system is empty is given by E in (5.83) and its value is found from (5.82) by applying L'Hopital's rule in the same manner.

$$E = \left\{ 1 - \sum_{k=1}^{n} \left(\rho_k \sum_{i=k}^{n} \beta_i \right) \right\} \Big/ \sum_{j=1}^{n} j\beta_j. \tag{5.90}$$

From this, we see that the condition for nonsaturation is

$$1 - \sum_{k=1}^{n} \left(\rho_k \sum_{i=k}^{n} \beta_i \right) > 0. \tag{5.91}$$

Here, we shall derive some relation between the generating functions in order to calculate the mean number of customers. Using Rouche's theorem, we can show that the denominator of the right-hand side of $U_l(x,y)$ $(1 \leq l \leq n)$ in (5.82), for each y in the domain $|y| \leq 1$, has just one root in the domain $|x| < 1$. Denote this root by $\delta(y)$. When $x = \delta(y)$, that is,

$$x = \{\sum_{j=1}^{n} \beta_j^{(1)} y^{1-\delta_{1,j}}\} P_1(x), \tag{5.92}$$

the numerator of the right-hand side of $U_l(x,y)$ $(1 \le l \le n)$ in (5.82) must be zero, since $|U_l(x,y)| \le 1$ $(1 \le l \le n)$ in the domain $|x| \le 1, |y| \le 1$. Thus, we get

$$\sum_{m=1}^{n} U_m(0,y) = \sum_{m=2}^{n-1} V_m^{(m)}(\delta(y),0,0) + W(\delta(y),0) + (\delta(y)-1)E + U_1(0,0). \qquad (5.93)$$

A similar discussion is also followed in the cases of $V_l^{(k)}(x,y,z)$ $(2 \le k \le n-1, k \le l \le n)$ and $W(x,z)$. Thus, from (5.82), we have

$$\sum_{m=1}^{n} U_m(0, A_2(x,z)) = \sum_{m=2}^{n} V_m^{(2)}(x,0,z) + U_1(0,0),$$

$$\sum_{m=k-1}^{n} V_m^{(k-1)}(x,0,A_k(x,z)) = \sum_{m=k}^{n} V_m^{(k)}(x,0,z) + V_{k-1}^{(k-1)}(x,0,0), \qquad 3 \le k \le n-1,$$

$$\sum_{m=n-1}^{n} V_m^{(n-1)}(x,0,A_n(x,z)) = W(x,0) + V_{n-1}^{(n-1)}(x,0,0), \qquad (5.94)$$

where

$$A_k(x,z) = \{\sum_{j=k}^{n} \beta_j^{(k)} z^{1-\delta_{k,j}}\} P_k(x), \qquad 2 \le k \le n. \qquad (5.95)$$

Using the above relation and considering (5.85), we obtain

$$\sum_{m=2}^{n-1} V_m^{(m)}(x,0,0) + W(x,0) = \sum_{m=1}^{n} U_m(0,Q(x)) - U_1(0,0),$$

$$\sum_{m=n-1}^{n} V_m^{(n-1)}(x,0,z) + \sum_{m=2}^{n-2} V_m^{(m)}(x,0,0) = \sum_{m=1}^{n} U_m(0,R(x,z)) - U_1(0,0), \qquad (5.96)$$

where

$$Q(x) = \sum_{k=2}^{n} \beta_k^{(2)} \prod_{i=2}^{k} P_i(x),$$

$$R(x,z) = \sum_{k=2}^{n-1} \beta_k^{(2)} \prod_{i=2}^{k} P_i(x) + \beta_n^{(2)} z \prod_{i=2}^{n-1} P_i(x). \qquad (5.97)$$

Now, we have reached a position to calculate the expected number of customers. This system has $n(n+1)/2$ kinds of instants of departure from a stage by combining a stage and a type. The average number of customers in a stage just after each departure can be obtained in the same manner. Thus, our objective is limited to calculating $E[Q_{1,1}^{(1)}]$, the mean number of customers in stage 1 just after a customer of type 1 has completed service in stage 1 and departs from stage 1 and $E[Q_{n,n}^{(1)}]+E[Q_{n,n}^{(n)}]$, the mean number of customers in the system just after a customer of type n has completed service in stage n and departs from stage n.

We have

$$E[Q_{1,1}^{(1)}] = \frac{1}{U_1(1,1)}\frac{\partial}{\partial x}U_1(x,y)|_{x=y=1},$$

$$\begin{aligned} Q_n &= E[Q_{n,n}^{(1)}]+E[Q_{n,n}^{(n)}] \\ &= \frac{1}{W(1,1)}\left[\frac{\partial}{\partial x}W(x,z)|_{x=z=1}+\frac{\partial}{\partial z}W(x,z)|_{x=z=1}\right]. \end{aligned} \tag{5.98}$$

The partial derivatives of $U_1(x,y)$ and $W(x,z)$ are obtained from (5.82). The limits of these derivatives as x, y and z tend to one are determined; however, L'Hopital's rule can be applied in order to resolve the indeterminacy. The first and second derivatives of some unknowns that we need to calculate (5.98) can be obtained from (5.93), (5.96) and (5.97). The derivatives of $\delta(y)$ are found, by implicit differentiation, from

$$\delta(y) = \{\sum_{j=1}^{n}\beta_j^{(1)}y^{1-\delta_{1,j}}\}P_1(\delta(y)). \tag{5.99}$$

We get the following results:

$$E[Q_{1,1}^{(1)}] = \frac{(1-\rho_1)\{P_1''(1) + b_n + 2\rho_1(1-\rho_1)\}}{2(1-\rho_1-a_n)(1-\rho_1+a_n)}, \tag{5.100}$$

$$\begin{aligned} Q_n &= \rho_n + \Big[(2a_{n-1} + \beta_n\rho_n + \beta_n)\{P_1''(1) + b_{n-1} + \beta_n P_n''(1) \\ &\quad + 2\rho_1(1-\rho_1)\} + 2\{(1-\rho_1-a_{n-1})(1-\rho_1+a_{n-1}) \\ &\quad + \beta_n^2\rho_n\}\left(\sum_{i=2}^{n-1}\rho_i\right)\Big]/\{2(1-\rho_1-a_n)(1-\rho_1+a_n)\}, \end{aligned} \tag{5.101}$$

where

$$a_m = \sum_{k=2}^{m}\left(\rho_k\sum_{i=k}^{n}\beta_i\right),$$

$$b_m = \sum_{k=2}^{m}\left\{P_k''(1)\sum_{i=k}^{n}\beta_i\right\} + 2\sum_{k=3}^{m}\left(\sum_{i=k}^{n}\beta_i\right)\left(\sum_{j=2}^{k-1}\rho_j\right)\rho_k. \tag{5.102}$$

Let

$$\rho_k = P_k''(1) = 0, \qquad 2 \le k \le n, \tag{5.103}$$

then (5.100) yields

$$\rho_1 + \frac{P_1''(1)}{2(1-\rho_1)} \tag{5.104}$$

coincides with the mean number of customers in the $M/G/1(\infty)$ system. Also, let

$$\beta_n = 1 \quad \text{and} \quad \beta_k = 0, \qquad 1 \le k \le n-1, \tag{5.105}$$

then (5.101) yields

$$\begin{aligned} \hat{Q}_n &= \rho_n + \Big[(2\hat{a}_{n-1} + \rho_n + 1)\{P_1''(1) + \hat{b}_{n-1} + P_n''(1) \\ &\quad + 2\rho_1(1-\rho_1)\} + 2\{(1-\rho_1-\hat{a}_{n-1})(1-\rho_1+\hat{a}_{n-1}) \\ &\quad + \rho_n\}\left(\sum_{i=2}^{n-1}\rho_i\right)\Big]/\{2(1-\rho_1-\hat{a}_n)(1-\rho_1+\hat{a}_n)\}, \end{aligned} \tag{5.106}$$

where

$$\hat{a}_m = \sum_{k=2}^{m} \rho_k,$$

$$\hat{b}_m = \sum_{k=2}^{m} P_k''(1) + 2\sum_{k=3}^{m}\left(\rho_k \sum_{i=2}^{k-1} \rho_i\right), \tag{5.107}$$

which gives the mean number of customers in the system just after a departure from stage n in the case that each customer receives service in all stages.

Let $n = 2$ in (5.106), then

$$\hat{Q}_2 = \rho_2 + \frac{(\rho_2 + 1)\{P_1''(1) + P_2''(1) + 2\rho_1(1 - \rho_1)\}}{2(1 - \rho_1 - \rho_2)(1 - \rho_1 + \rho_2)}, \tag{5.108}$$

which agrees with the result of Taube-Netto [100].

The steady state condition of the system in which each customer receives service in all stages, from (5.91), is

$$1 - \sum_{k=1}^{n} \rho_k > 0. \tag{5.109}$$

It is easily seen that (5.109) implies (5.91). If (5.109) holds and total arrival rate λ has the same value in both systems, then it can be proved that Q_n is smaller than $\hat{Q}_n$. This fact is natural since some customers leave the system without receiving service in all stages in our model.

5.5. Concavity or Convexity of Performance Measures of Two-Stage Finite Tandem Queueing System

5.5.1. Model and Steady State Probabilities

Fig. 5-4 shows a tandem queueing system with two stages numbered $1, 2$. Stage j has a single server and the service times are independent and exponentially distributed with mean $1/\mu_j$ $(j = 1,2)$. The buffer storage capacity between stages 1 and 2 is $N-1(N \geq 2)$, consequently there are more than one buffer space between two stages. No waiting is allowed in front of stage 1 . Therefore, if stage 1 is occupied, then an arriving customer cannot enter stage 1 . The successive customers arrive according to a Poisson law with parameter λ . Customers require service from all stages in the order $1, 2$ and the service to a customer at stage 1 is initiated even if the intermediate buffer is full. In this case, a server at stage 1 is blocked only if the customer it served cannot be advanced to stage 2 . This is called manufacturing blocking (or production blocking). Arrivals at stage 1 when a server is blocked are turned away. Service to a customer at any stage, once initiated, is completed without interruptions.

We now define the following steady state probabilities :

$p_{i,j}(0 \leq i \leq 1, 0 \leq j \leq N)$: i customers are in stage 1, and j customers are in stage 2 ;

$p_{b,N}$: a server at stage 1 is blocked ;

p_0 : no customers are in the system.

The steady state difference equations may be constructed in the usual manner.

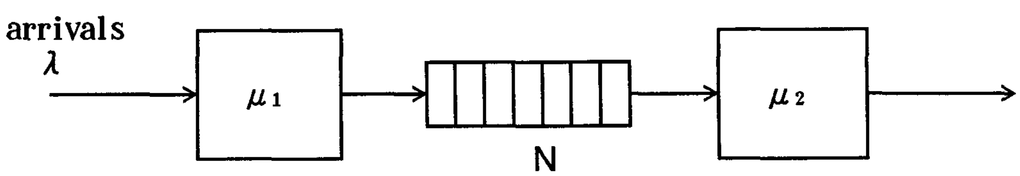

Fig. 5-4. Two-stage tandem queueing system with finite intermediate buffers.

They are

$$
\begin{aligned}
(\lambda+\mu_2)p_{0,i} &= \mu_1 p_{1,i-1}+\mu_2 p_{0,i+1}, \qquad & 1\le i\le N-1,\\
(\mu_1+\mu_2)p_{1,i} &= \lambda p_{0,i}+\mu_2 p_{1,i+1}, \qquad & 1\le i\le N-1,\\
\lambda p_0 &= \mu_2 p_{0,1}, &\\
(\lambda+\mu_2)p_{0,N} &= \mu_1 p_{1,N-1}+\mu_2 p_{b,N}, &\\
\mu_1 p_{1,0} &= \lambda p_0+\mu_2 p_{1,1}, &\\
(\mu_1+\mu_2)p_{1,N} &= \lambda p_{0,N}, &\\
\mu_2 p_{b,N} &= \mu_1 p_{1,N}. & \qquad (5.110)
\end{aligned}
$$

Also the normalization condition yields

$$
p_0+\sum_{i=1}^{N} p_{0,i}+\sum_{i=0}^{N} p_{1,i}+p_{b,N}=1. \qquad (5.111)
$$

We have put

$$
\rho=\frac{\lambda}{\mu_1},
$$

$$
\gamma=\frac{\mu_2}{\mu_1}, \qquad (5.112)
$$

and

$$
a_0=1,\;\; a_i=(\rho+\gamma)b_i-\gamma\rho b_{i-1}, \qquad i\ge 1,
$$

$$
b_0=\frac{1}{\gamma},\; b_1=1+\gamma,\; b_i=(1+\gamma)a_{i-1}-\gamma\rho a_{i-2}, \qquad i\ge 2. \qquad (5.113)
$$

Thus, solving (5.110) and (5.111), we can determine the steady state probabilities for $N \geq 2$:

$$
\begin{aligned}
p_{0,i} &= \gamma\rho^i b_{N-i+1}/D, && 1 \leq i \leq N-1, \\
p_{1,i} &= \gamma\rho^{i+1} a_{N-i}/D, && 0 \leq i \leq N-1, \\
p_{0,N} &= \gamma(1+\gamma)\rho^N/D, \\
p_{1,N} &= \gamma\rho^{N+1}/D, \\
p_{b,N} &= \rho^{N+1}/D, \\
p_0 &= \gamma^2 b_N/D,
\end{aligned}
\tag{5.114}
$$

where

$$D = \frac{\gamma(\rho+1)a_{N+1} - \gamma^3(\rho+1)a_N - \rho^{N+2}}{\{(\gamma-1)\rho+\gamma\}}, \tag{5.115}$$

and

$$a_i = \gamma^i \sum_{k=0}^{[i/2]} (-1)^k \binom{i-k}{k} \rho^k (\rho+\gamma+1)^{i-2k}, \quad i \geq 1. \tag{5.116}$$

The above explicit results of the steady state probabilities owe to the recursive relation (5.113) about a_i and b_i. By mathematical induction, we can obtain the following lemma, which will be useful in later discussion.

Lemma 5.1. *If the recursive relation (5.113) holds, then we have for $i \geq 1$ and $\gamma \geq 1$,*

$$(i) \quad a_{i+1} = \gamma(\rho+\gamma+1)a_i - \gamma^2\rho a_{i-1}. \tag{5.117}$$

$$(ii) \quad a_{i-1} > 0, b_{i-1} > 0, a_i > \rho a_{i-1}. \tag{5.118}$$

$$(iii) \quad a_i^2 > a_{i-1}a_{i+1}. \tag{5.119}$$

5.5.2. Concavity of Mean Number of Busy Service Stations

From (5.114), the mean number of busy stations is derived as

$$B = \sum_{i=1}^{N} p_{0,i} + p_{1,0} + p_{b,N} + 2\,(\sum_{i=1}^{N} p_{1,i})$$

$$= \frac{(\gamma+1)(\rho a_{N+1} - \gamma^2 \rho a_N - \rho^{N+2})}{\{(\gamma-1)\rho+\gamma\}\cdot D}. \tag{5.120}$$

Hence,

$$B \rightarrow \begin{cases} \dfrac{\gamma+1}{\gamma}\rho, & \rho \ll 1, \\ \dfrac{(\gamma+1)(a_{N+1} - \gamma^2 a_N - 1)}{2\gamma a_{N+1} - 2\gamma^3 a_N - 1}, & \rho = 1, \\ \dfrac{(\gamma+1)(\gamma^{N+1}-1)}{\gamma^{N+2}-1}, & \rho \gg 1. \end{cases} \tag{5.121}$$

Substituting $N = 2$ and $\gamma = 1$ in (5.120) and (5.121), we get the results given by Morse [63].

On the other hand, the utilization per one station is

$$\frac{B}{\frac{1}{\mu_1} + \frac{1}{\mu_2}}, \tag{5.122}$$

and comparing the coefficients of ρ^{N+2} in both the numerator and the denominator of (5.122), the maximum allowable ρ for steady state existence in this system is

$$\rho_{max} = \lim_{\rho\to\infty} \frac{\mu_1\mu_2}{\mu_1+\mu_2}\cdot B = \frac{\mu_2(\mu_1^{N+1} - \mu_2^{N+1})}{\mu_1^{N+2} - \mu_2^{N+2}}, \tag{5.123}$$

and for $\mu_1 = \mu_2$ this reduces to

$$\rho_{max} = \frac{N+1}{N+2}. \tag{5.124}$$

Results in (5.123) and (5.124) on ρ_{max} coincide with ones given by Hunt [37].

Now let us return to a discussion of the concavity.

Theorem 5.1. *Let $B(N) = F(N)/G(N)$ be a function of the discrete variable N, then $B(N)$ is strictly increasing and concave for $\gamma \geq 1$.*

Proof. Proof is given by elementary but a little troublesome calculation as follows.

From (ii) in Lemma 5.1,

$$G(N+2), G(N+1), G(N) > 0. \tag{5.125}$$

Now, using the recursive formula (i) in Lemma 5.1,

$$\Delta B(N) = F(N+1)/G(N+1) - F(N)/G(N)$$

$$= \frac{(\gamma+1)\{(\gamma-1)\rho+\gamma\}^2 \rho^{N+2} a_{N+1}}{G(N+1)\cdot G(N)} > 0, \tag{5.126}$$

and hence

$$\Delta^2 B(N) = (\gamma+1)\{(\gamma-1)\rho+\gamma\}^2 \{\frac{\rho^{N+3}a_{N+2}}{G(N+2)\cdot G(N+1)} - \frac{\rho^{N+2}a_{N+1}}{G(N+1)\cdot G(N)}\}$$

$$= \frac{-\rho^{N+2}(\gamma+1)\{(\gamma-1)\rho+\gamma\}^2}{G(N+2)\cdot G(N+1)\cdot G(N)}$$

$$\cdot [\gamma^2(\rho+1)\{(\gamma-1)\rho+\gamma\}a_{N+1}b_{N+1} + \gamma^2\rho^2(\rho+1)(\gamma-1)a_{N+1}^2$$

$$+ \gamma^3\rho(\rho+1)a_N a_{N+2} + \rho^{N+3}(a_{N+2} - \rho a_{N+1})]. \tag{5.127}$$

From (ii) in Lemma 5.1, if $\gamma \geq 1$, then

$$\Delta^2 B(N) < 0. \quad \square \tag{5.128}$$

5.5.3. Convexity of Rate of Loss Calls

Similarly from (5.114), we have the rate of loss calls (i.e., the equilibrium probability that an arrival is lost) ;

$$P_c = p_{b,N} + \sum_{i=0}^{N} p_{1,i}$$

$$= \frac{\gamma\rho a_{N+1} - \gamma^3 \rho a_N - \rho^{N+2} + \gamma\rho^{N+1}}{\{(\gamma-1)\rho+\gamma\}\cdot D}. \tag{5.129}$$

Hence,

$$P_c \to \begin{cases} \rho - \rho^2, & \rho \ll 1, \\ \dfrac{\gamma a_{N+1} - \gamma^3 a_N + \gamma - 1}{2\gamma a_{N+1} - 2\gamma^3 a_N - 1}, & \rho = 1, \\ 1 - \dfrac{\gamma(\gamma^{N+1}-1)}{(\gamma^{N+2}-1)\rho}, & \rho \gg 1. \end{cases} \tag{5.130}$$

Substituting $N = 2$ and $\gamma = 1$, (5.129) and (5.130) yield to the results given by Morse [63].

Then, we shall show the following convexity property for P_c.

Theorem 5.2. *Let P_c be a function of the discrete variable N, then P_c is strictly decreasing and convex for $\gamma \geq 1$.*

Proof. Some calculations yield

$$\Delta P_c(N) = -\frac{\gamma}{(\gamma+1)\rho}\cdot \Delta B(N), \tag{5.131}$$

where $\Delta B(N)$ is given by (5.126). Since $\Delta B(N) > 0$ and $\Delta^2 B(N) < 0$ for $\gamma \geq 1$ from Theorem 5.1, we get for $\gamma \geq 1$

$$\Delta P_c(N) < 0 \text{ and } \Delta^2 P_c(N) > 0. \quad \square \tag{5.132}$$

Here we shall end this section to demonstrate a similar property for other performance measure. It is the mean number of customers given by calculating from (5.114)as follows.

$$L = \sum_{i=1}^{N} i p_{0,i} + \sum_{i=0}^{N} (i+1) p_{1,i} + (N+1) p_{b,N}$$

$$= [\gamma\rho(\gamma\rho+1)a_{N+1} - \gamma^2(2\gamma-1)\rho^2 a_N$$

$$- \{(\gamma-1)N + (\gamma^2+\gamma-1)\}\rho^{N+3} - \gamma(N+2)\rho^{N+2}]/[\{(\gamma-1)\rho+\gamma\}^2 \cdot D]. \tag{5.133}$$

Hence,

$$L \to \begin{cases} \dfrac{\gamma+1}{\gamma}\rho, & \rho \ll 1, \\ \dfrac{\gamma(\gamma+1)a_{N+1} - \gamma^2(2\gamma-1)a_N - \gamma^2 - (2N+3)\gamma + N + 1}{(2\gamma-1)(2\gamma a_{N+1} - 2\gamma^3 a_N - 1)}, & \rho = 1, \\ \dfrac{\gamma^{N+3} - \gamma^2 - (N+1)\gamma + N + 1}{(\gamma-1)(\gamma^{N+2}-1)}, & \rho \gg 1. \end{cases} \tag{5.134}$$

Substituting $N = 2$ and $\gamma = 1$ in (5.133) and (5.134), we get the results given by Morse [63].

The equation (5.133) is too complicated to handle in a straightforward manner. But we find the following property after a troublesome calculations.

Theorem 5.3. *Let $L(N)$ be a function of the discrete variable N, then $L(N)$ is strictly increasing for $\gamma \geq 1$ and $\rho < 1$.*

Proof. In a manner similar to Theorem 5.1 and Theorem 5.2, we calculate $\Delta L(N)$. And if we use (i) and (iii) in Lemma 5.1, $N \geq 1$ and $\rho < \rho_{max} < 1$, then we can derive

$$\Delta L(N) > 0. \quad \square \tag{5.135}$$

5.6. Conclusion

We consider a tandem queueing system with several types of customers in which service distribution is exponential or general in this chapter. First, both for a two-stage ordinary and commutative tandem queueing systems with two types of customers and independent service times, we obtained the mean queue length by a generating function approach and showed that when a customer departs from the system without receiving the second service does not arrival, i.e., the arrival rate of a customer of type 1 equals zero, then the consequent result agrees with the results which have already been derived. Secondly, for a multi-stage tandem queueing system attended by a single server with general service times, zero switchover time and a exhaustive switching rule, we calculated the mean number of customers in the system just after a customer of type n by utilizing the relation between the generating functions and showed that it is lower than in the case of a model with one type of customers if the steady state condition holds. Finally, for a two-stage tandem queueing system with independent service times, no queues ahead of the first stage and finite intermediate buffers, we obtained by an elementary calculation the mean number of busy service stations, the rate of loss calls and the mean number of customers in the system and established the monotonicity and concavity of the mean number of busy service stations, the monotonicity and convexity of the rate of loss calls and the monotonicity of the mean number of customers in the system with respect to intermediate buffer spaces.

Chapter 6.

MULTISERVER QUEUEING SYSTEMS WITH ADDITIONAL SERVICE CHANNELS

6.1. Introduction

We described in Chapter 1 that there is one of ways building models which do not include the assumption of independence between arrival and service patterns: service parameters are managed to vary with the present number of customers. This implies a semi-Markov process for the system state. However, in practice, many situations are found in which service parameters, when assuming that the number of channels is a constant, do not vary with the present number of customers, but the number of channels vary with the present number of customers when assuming that service parameters are constant. A typical example would be a tool crib that has two clerks but usually keeps only one at the service window. When arrivals happen to occur close together and a queue of undesirable length develops, the second clerk leaves his other work and helps at the window.

Romani [91] considered the process in which the queue is never allowed to grow beyond a specified length M. Whenever M customers are in the queue and there is an arrival, the number of channel is increased by one. There is no limit to the number of channels that may be added. When there are no customers in the queue and service are completed, the channels that are empties are cancelled.

Morder and Phillips [62] studied the process in which the number of channels is limited and the queue is unlimited. The number of manned channels increases from a fixed minimum number, σ, when the queue reaches a given length N. When the maximum number of channels $S(>\sigma)$ are operating, no further increases are possible and the queue is unbounded. Channels are cancelled when the number of customers in the queue drops to $\nu(0 \leq \nu \leq N-2)$ and a service is completed.

Murari [65] considered the queueing problem with variable number of channels assuming that when the queue length increases to some undesirable number m_1, then another channel is called for its help, and if in spite of two channels operating the queue length increases to some undesirable number $m_2(> m_1)$, then the third channel is called for their help and so on, and in all cases when there is a demand of an additional channel it is made available instantaneously and is cancelled at the termination of service if there are no customers waiting, with exception of one channel which remains open at all times.

Singh [95] studied the steady state behavior of the M/M/1(M) system assuming that when the queue length reaches M, then a search for an additional service facility is started and the availability time is exponentially distributed.

In this chapter, we deal with a modification of Murari's model including the search of an additional channel. In Section 6.2, we describe our model fully and give a explicit expression that formulate all of the steady state probabilities. In Section 6.3, we have a summary of this chapter.

6.2. Analysis of Model

6.2.1. Model

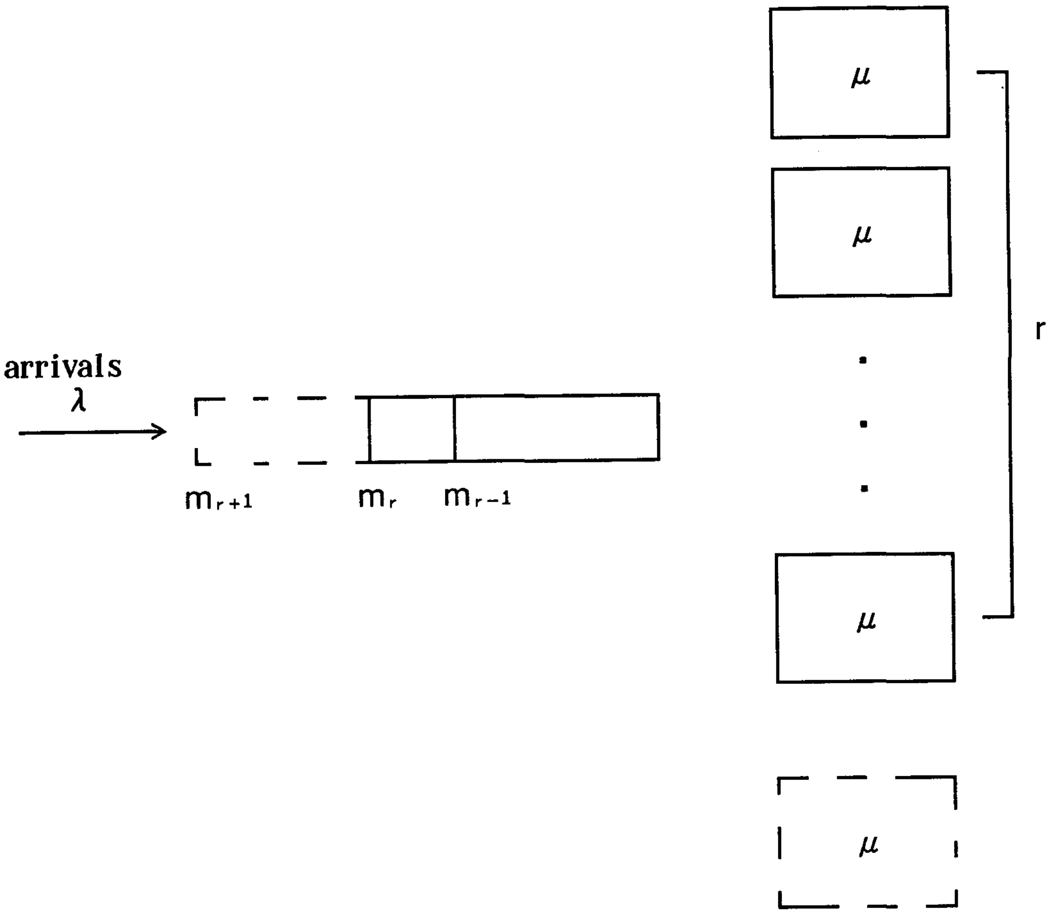

Fig. 6-1. Multiserver queueing system with additional service channels.

Fig. 6-1 shows a multiserver queueing system with additional service channels, which is the modification of the model in Murari [65] as follows. When the queue length increases to some undesirable number m_1, then a search for another channel is started and the availability time of another channel is exponentially distributed. The search continues till the queue length reduces to m_0 ($< m_1$). If the search is fulfilled, another channel is available and it takes in the head of the queue for service instantaneously. If in spite of two service channels operating the queue length increases to some undesirable number m_2 ($> m_1$), then a search for the third channel is started and so on.

Thus, we study the steady state behavior of a queueing system under the following assumptions:

(1)Customers arrive at a service facility in a Poisson stream with mean rate λ and form a queue.

(2)The queue discipline is first-come first-served.

(3)The service facility consists of one regular service channel and c additional service channels (abbreviated as a.s.c.'s). The regular service channel is always at the disposal of customers, irrespective of the queue length. However, the number of operating a.s.c.'s at any instant is dependent of the queue length:

(3.1)If the queue length is zero and the regular service channel is idle, it takes in a unit for service instantaneously on arrival.

(3.2)If the queue length is m_r $(1 \leq r \leq c)$, and $0 \leq m_1 \leq m_2 \leq \cdots \leq m_c$ and the number of operating a.s.c.'s is $r-1$, then a search for the next a.s.c. is started and the availability time (the time when the next a.s.c. is made available, measured from the instant that a search is started) is exponentially distributed with parameter V. The search continues till the queue length reduces to m_{r-1} $(1 \leq r \leq c)$. If during the small interval $[t, t+\Delta t)$ the search is fulfilled, that is , the next a.s.c. is available, the head of the queue gets service from the next a.s.c..

(3.3)If the queue length is zero at the completion of service at a service channel (regular or additional) then one of the operating a.s.c.'s, if there is any, stops operating.

(4)All service channels have the same exponential service time distribution with mean service time $1/\mu$.

(5)If at any instant the queue length is m_r or K ($> m_c$) and the number of operating a.s.c.'s is $r-1$ or equal to c, respectively, then the arriving customers will be considered lost for the system.

(6)The stochastic processes involved are independent of each other.

We define the following steady state probabilities:

Q : the queue length is zero and no service channel is operating;

$p^{(A)}_{r,n}$: the queue length is n, the regular service channel is busy, the number of operating a.s.c.'s is r ($0 \leq r \leq c-1$) and a search for the next a.s.c. is in progress;

$p^{(B)}_{r,n}$: the queue length is n, the regular service channel is busy, the number of operating a.s.c.'s is r ($0 \leq r \leq c$) and no search for the next a.s.c. is in progress.

Thus by the assumptions imposed on the system

$$p^{(A)}_{r,n} = 0, \qquad 0 \leq n \leq m_r, 0 \leq r \leq c-1,$$

$$p^{(A)}_{c,n} = 0, \qquad 0 \leq n \leq K,$$

$$p^{(B)}_{r,m_r+1} = 0, \qquad 0 \leq r \leq c-1. \tag{6.1}$$

6.2.2. Steady State Probabilities

The steady state equations may be constructed in the usual manner. For $0 \leq r \leq c-1$, we have

$$\{\lambda+(r+1)\mu+V\}p^{(A)}_{r,1+m_r}=(r+1)\mu p^{(A)}_{r,2+m_r},$$

$$\{\lambda+(r+1)\mu+V\}p^{(A)}_{r,n}=\lambda p^{(A)}_{r,n-1}+(r+1)\mu p^{(A)}_{r,n+1},\quad 2+m_r\le n\le -1+m_{r+1},$$

$$\{(r+1)\mu+V\}p^{(A)}_{r,m_{r+1}}=\lambda p^{(A)}_{r,-1+m_{r+1}}+\lambda p^{(B)}_{r,-1+m_{r+1}},$$

$$\lambda Q=\mu p^{(B)}_{0,0},$$

$$(\lambda+\mu)p^{(B)}_{0,0}=\lambda Q+\mu p^{(B)}_{0,1}+2\mu p^{(B)}_{1,0},$$

$$(\lambda+\mu)p^{(B)}_{0,n}=\lambda p^{(B)}_{0,n-1}+\mu p^{(B)}_{0,n+1},\qquad 1\le n\le -2+m_1, n\ne m_0,$$

$$(\lambda+\mu)p^{(B)}_{0,m_0}=\lambda p^{(B)}_{0,-1+m_0}+\mu p^{(B)}_{0,1+m_0}+\mu p^{(A)}_{0,1+m_0},$$

$$(\lambda+\mu)p^{(B)}_{0,-1+m_1}=\lambda p^{(B)}_{0,-2+m_1}$$

and for $1\le r\le c-1$,

$$\{\lambda+(r+1)\mu\}p^{(B)}_{r,0}=(r+1)\mu p^{(B)}_{r,1}+(r+2)\mu p^{(B)}_{r+1,0},$$

$$\{\lambda+(r+1)\mu\}p^{(B)}_{r,n}=\lambda p^{(B)}_{r,n-1}+(r+1)\mu p^{(B)}_{r,n+1},$$

$$1\le n\le -1+m_{r-1}, 1+m_r\le n\le -2+m_{r+1},$$

$$\{\lambda+(r+1)\mu\}p^{(B)}_{r,n}=\lambda p^{(B)}_{r,n-1}+(r+1)\mu p^{(B)}_{r,n+1}+Vp^{(A)}_{r-1,n+1},$$

$$m_{r-1}\le n\le -1+m_r,$$

$$\{\lambda + (r+1)\mu\}p^{(B)}_{r,m_r} = \lambda p^{(B)}_{r,-1+m_r} + (r+1)\mu p^{(B)}_{r,1+m_r} + (r+1)\mu p^{(A)}_{r,1+m_r},$$

$$\{\lambda + (r+1)\mu\}p^{(B)}_{r,-1+m_{r+1}} = \lambda p^{(B)}_{r,-2+m_{r+1}},$$

and

$$\{\lambda + (c+1)\mu\}p^{(B)}_{c,0} = (c+1)\mu p^{(B)}_{c,1},$$

$$\{\lambda + (c+1)\mu\}p^{(B)}_{c,n} = \lambda p^{(B)}_{c,n-1} + (c+1)\mu p^{(B)}_{c,n+1},$$

$$1 \le n \le -1 + m_{c-1}, m_c \le n \le K-1,$$

$$\{\lambda + (c+1)\mu\}p^{(B)}_{c,n} = \lambda p^{(B)}_{c,n-1} + (c+1)\mu p^{(B)}_{c,n+1} + Vp_{c-1,n+1},$$

$$m_{c-1} \le n - 1 + m_c,$$

$$(c+1)\mu p^{(B)}_{c,K} = \lambda p^{(B)}_{c,K-1}. \qquad (6.2)$$

We introduce the following generating functions:

$$P^{(A)}_r(z) = \sum_{n=1+m_r}^{m_{r+1}} z^n p^{(A)}_{r,n}, \qquad 0 \le r \le c-1,$$

$$P^{(B)}_r(z) = \sum_{n=0}^{-1+m_{r+1}} z^n p^{(B)}_{r,n}, \qquad 0 \le r \le c-1,$$

$$P^{(B)}_c(z) = \sum_{n=0}^{K} z^n p^{(B)}_{c,n}. \qquad (6.3)$$

Multiplying the equations (6.2) by appropriate powers of z and using (6.3), we have

$$P_r^{(A)}(z) = \frac{\lambda(1-z)z^{1+m_{r+1}}p_{r,m_{r+1}}^{(A)} + \lambda z^{1+m_{r+1}}p_{r,-1+m_{r+1}}^{(B)} - (r+1)\mu z^{1+m_r}p_{r,1+m_r}^{(A)}}{\{\lambda+(r+1)\mu+V\}z - \lambda z^2 - (r+1)\mu},$$

$$0 \le r \le c-1,$$

$$P_0^{(B)}(z) = \frac{\lambda z Q + 2\mu z p_{1,0}^{(B)} + \mu z^{1+m_0}p_{0,1+m_0}^{(A)} - \mu p_{0,0}^{(B)} - \lambda z^{1+m_1}p_{0,-1+m_1}^{(B)}}{(\lambda+\mu)z - \lambda z^2 - \mu},$$

$$P_r^{(B)}(z) = \{(r+2)\mu z p_{r+1,0}^{(B)} + (r+1)\mu z^{1+m_r}p_{r,1+m_r}^{(A)} + VP_{r-1}^{(A)}(z) - (r+1)\mu p_{r,0}^{(B)}$$

$$- \lambda z^{1+m_{r+1}}p_{r,-1+m_{r+1}}\}/[\{\lambda+(r+1)\mu\}z - \lambda z^2 - (r+1)\mu],$$

$$1 \le r \le c-1,$$

$$P_c^{(B)}(z) = \frac{\lambda(1-z)z^{K+1}p_{c,K}^{(B)} + VP_{c-1}^{(A)}(z) - (c+1)\mu p_{c,0}^{(B)}}{\{\lambda+(c+1)\mu\}z - \lambda z^2 - (c+1)\mu}. \tag{6.4}$$

Since $P_r^{(A)}(z)$ $(0 \le r \le c-1)$ and $P_r^{(B)}(z)$ $(0 \le r \le c)$ are regular on z-plane, the zeros of the numerator and the denominator of the right-hand side of $P_r^{(A)}(z)$ and $P_r^{(B)}(z)$, which are given in (6.4), must coincide with each other, giving $4c+2$ equations. Solving this set of equations together with the steady state equation

$$\lambda Q = \mu p_{0,0}^{(B)}, \tag{6.5}$$

we can determine the $4c+3$ unknowns occuring on the numerator of the right-hand side of (6.4) as follows.

For $0 \leq r \leq c-1$,

$$p^{(A)}_{r,1+m_r} = \Big\{\Big(\prod_{i=1}^{r} g_i\Big)Q\Big\} \Big/ \{(r+1)\beta_0(b_0+c_0-b_0d_0-1)\},$$

$$p^{(A)}_{r,m_{r+1}} = (r+1)\beta_0 a_r p^{(A)}_{r,1+m_r},$$

$$p^{(B)}_{r,-1+m_{r+1}} = (r+1)\beta_0 b_r p^{(A)}_{r,1+m_r},$$

$$p^{(B)}_{r+1,0} = \{(r+1)(b_r-1)p^{(A)}_{r,1+m_r}\}/(r+2),$$

and

$$p^{(B)}_{0,0} = Q/\beta_0,$$

$$p^{(B)}_{c,K} = \Big\{(c_c-b_cd_c)\Big(\prod_{i=1}^{c} g_i\Big)Q\Big\} \Big/ \{\beta_c^{K+1}(b_0+c_0-b_0d_0-1)\}, \tag{6.6}$$

where

$$a_r = \Big\{\sum_{i=0}^{-1-m_r+m_{r+1}} (\gamma_{r+1}/\kappa_{r+1})^i\Big\} \Big/ (\kappa_{r+1}\gamma_{r+1}^{-m_r+m_{r+1}}), \qquad 0 \leq r \leq c-1,$$

$$b_r = \Big\{\kappa_{r+1} \sum_{i=0}^{-m_r+m_{r+1}} (\gamma_{r+1}/\kappa_{r+1})^i$$

$$- \sum_{i=0}^{-1-m_r+m_{r+1}} (\gamma_{r+1}/\kappa_{r+1})^i\Big\} \Big/ (\kappa_{r+1}\gamma_{r+1}^{-m_r+m_{r+1}}), \quad 0 \leq r \leq c-1,$$

$$c_0 = 1 - \sum_{i=1}^{m_0} \beta_0^i, \qquad d_0 = 1 - \sum_{i=1}^{m_1} \beta_0^i,$$

$$c_r = 1 - \sum_{i=0}^{m_r} \beta_r^i, \qquad d_r = 1 - \sum_{i=0}^{m_{r+1}} \beta_r^i, \qquad 1 \le r \le c-1,$$

$$e_{r-1} = -(V/\lambda) \sum_{i=1}^{m_{r-1}} \beta_r^i - \beta_r + \gamma_r \kappa_r, \qquad 1 \le r \le c,$$

$$f_{r-1} = -(V/\lambda) \sum_{i=1}^{m_r} \beta_r^i - \beta_r + \gamma_r \kappa_r, \qquad 1 \le r \le c,$$

$$g_r = [\{(V/\lambda) a_{r-1} \beta_r^{1+m_r} + b_{r-1} f_{r-1} - e_{r-1}\}$$

$$\cdot (\kappa_r - \beta_r)(b_r d_r - c_r)]/(\beta_r - \gamma_r), \qquad 1 \le r \le c,$$

$$\beta_r = \{(r+1)\mu\}/\lambda, \qquad 0 \le r \le c,$$

$$\gamma_r = [\lambda + r\mu + V + \{(\lambda + r\mu + V)^2 - 4r\lambda\mu\}^{\frac{1}{2}}]/(2\lambda), \qquad 1 \le r \le c,$$

$$\kappa_r = [\lambda + r\mu + V - \{(\lambda + r\mu + V)^2 - 4r\lambda\mu\}^{\frac{1}{2}}]/(2\lambda), \qquad 1 \le r \le c, \tag{6.7}$$

and Q can be determined by the normalization condition

$$Q + \sum_{r=0}^{c-1} P_r^{(A)}(1) + \sum_{r=0}^{c} P_r^{(B)}(1) = 1. \tag{6.8}$$

Comparing the coefficient of z^n on both sides of (6.4) and using (6.6), we have, for $c \ge 1$,

$$p_{r,n}^{(A)} = \left[\left\{ \sum_{i=0}^{n-1-m_r} \left(\gamma_{r+1}/\kappa_{r+1}\right)^i \right\} p_{r,1+m_r}^{(A)} \right] \Big/ \gamma_{r+1}^{n-1-m_r},$$

$$0 \le r \le c-1, 2+m_r \le n \le m_{r+1},$$

$$p_{0,n}^{(B)} = b_0\Big(\sum_{i=1}^{-n+m_1} \beta_0^i - \sum_{i=1}^{-n+m_0} \beta_0^i \Big) p_{0,1+m_0}^{(A)}, \qquad 1 \le n \le -1+m_0,$$

$$p_{0,n}^{(B)} = b_0 p_{0,1+m_0}^{(A)} \sum_{i=1}^{-n+m_1} \beta_0^i, \qquad m_0 \le n \le -1+m_1,$$

$$p_{c,n}^{(B)} = \beta_c^{K-n} p_{c,K}^{(B)}, \qquad 1+m_c \le n \le K, \tag{6.9}$$

and specially, in the case $c = 1$ and $m_0 = 0$,

$$p_{0,1}^{(A)} = Q/\{\beta(b_0 - b_0 d_0 - 1)\},$$

$$p_{1,n}^{(B)} = p_{1,0}^{(B)}/\kappa_1^n + \sum_{i=1}^{n} \Big\{ \sum_{j=K+i-n-1}^{K} h_{n-i+1}^{(j)} \Big\} \Big/ \kappa_1^i, \quad 1 \le n \le m_1, \tag{6.10}$$

where

$$h_n^{(i)} = \Big\{ -(V/\lambda) b_0 \sum_{j=1}^{n+i-K} \gamma^{-j} + (1-b_0)(1-\kappa_1)\gamma_1^{K-n-i} \Big\} \beta_0 \beta_1^{i-K-1} p_{0,1}^{(A)},$$

$$h_n^{(K-n)} = (b_0 - 1)\kappa_1 \beta_1^{-n-1}. \tag{6.11}$$

The results due to Murari [65] can be obtained by exchanging $p_{r,n}^{(B)}$ $(0 \le r \le c-1, 0 \le n \le -1+m_{r+1}; r = c, 0 \le n \le K)$ and $-1+m_{r+1}$ $(0 \le r \le c-1)$ with $p_{r,n}$ and m_{r+1}, respectively, and letting V tend to infinity in our model.

6.3. Conclusion

We considered a multiserver queueing system with additional service channels in which service times is exponentially distributed. This model was a modification of Murari's model [65] including the search of additional channel like Singh [95]. We gave the explicit expression of the steady state probabilities by a generating function approach and showed that we can obtain the results due to Murari [65] by exchanging the definitions and letting a parameter V of the availability time tend to infinity.

Chapter 7.

CONCLUSION

In this thesis, we have discussed three types of queueing systems, i.e., correlated service times, several types of customers and additional service channels. Various explicit solutions have been provided for these three types of queueing systems. Here we summarize the obtained results and discuss the further direction of an analysis for queueing systems with correlated service times.

In Chapter 2 through 4, we have dealt with queueing systems with correlated service times whose distribution is the bivariate or multivariate exponential distribution of Marshall and Olkin discussed in Chapter 1.

Chapter 2 discussed a multiserver queueing system both for no jockeying and jockeying cases. For the no jockeying case, the steady state probabilities and the queueing and waiting time distribution have been derived and it has been proved that the consequent result agrees with the so-called Little's formula. For the jockeying case, an explicit solution of the rate matrix R has been derived by utilizing the results obtained in the no jockeying case and enabled us to calculate the steady state probability vector, the mean queue length and the mean waiting time.

Chapter 3 has discussed a two-stage ordinary tandem queueing system and a three-stage commutative tandem queueing system. For the ordinary tandem case, the mean number of customers in the system has been obtained and it has been shown analytically that it decreases as the correlation coefficient increases. For the commutative tandem case, the throughput rate of the system has been calculated by a matrix-geometric approach. Further it has been established that when the correlation coefficient is small, the throughput rate is lower in the case of an ordinary service operation and correlated service times than in the case of a commutative service operation and independent service times, but when the correlation coefficient is large, the reverse is true. Furthermore, this chapter has discussed the effect of correlated service times. It has been proved that in the case of a two-server parallel queueing system treated in Section 2.2, the mean number of customers is greater when service times are correlated than when service times are independent, but in the case of a two-stage ordinary tandem queueing system discussed in Section 3.2, the reverse is true, if the system can finish 2μ customers' service per unit time on the average.

Chapter 4 has discussed a interchangeable parallel two-stage tandem queueing system. The throughput rate of the system has been obtained by a matrix-geometric approach and it has been established that for a small correlation coefficient, the throughput rate is lower in the case of an ordinary service operation and correlated service times than in the case of an interchangeable service operation and independent service times, but for a large correlation coefficient, the reverse is true.

Chapter 5 has discussed tandem queueing systems with several types of customers for two cases of finite and infinite intermediate buffers. First, both for a two-stage ordinary and commutative tandem queueing system with no intermediate buffers, the mean queue length has been obtained by an elementary but troublesome calculations. Secondly, for a multi-stage tandem queueing system attended by a single server with infinite intermediate buffers, general service times, zero switchover time and a exhaustive switching rule, the mean number of customers has been calculated by utilizing the relation between the generating functions. Finally, for a two-stage tandem queueing system with finite intermediate buffers, the mean number of busy service stations, the rate of loss calls and the mean number of customers have been calculated by solving balance equations directly. And it has been proved that the mean number of service stations is strictly increasing and concave, the rate of loss calls is strictly decreasing and convex and the mean number of customers is strictly increasing with respect to intermediate buffer spaces.

Chapter 6 has discussed a multiserver queueing system with additional service channels. The explicit expression of the steady state probabilities has been derived by a generating function approach.

As mentioned in Section 1.3, the multivariate exponential distribution of Marshall and Olkin we adopted as the service distribution in Chapter 2 through Chapter 4 is no more than a multivariate phase type distribution. Many researchers have studied parallel or tandem queueing systems whose service distribution is phase type. But most of queueing systems with phase type service times are for the univariate cases, or although they are for the multivariate cases, the service times are assumed to be independent (see Neuts [71], and Gün and Makowski [30]). Therefore, the further development of the analysis for queueing systems with correlated service times treated in this thesis is expected to be done for a multivariate phase type distribution which includes a wide class of distribution.

References

1. I. J. B. F. Adan, J. Wessels and W. H. M. Zijm, Matrix-geometric analysis of the shortest queue problem with threshold jockeying, *Oper. Res. Lett.* **13** (1993), 107–112.

2. V. Anantharam and P. Tsoucas, Stochastic concavity of throughput in series of queues with finite buffers, *Adv. Appl. Prob.* **22** (1990), 761–763.

3. C. J. Ancker, Jr. and A. V. Gafarian, Queuing with multiple Poisson inputs and exponential service times, *Opns. Res.* **9** (1961), 321–327.

4. D. Assaf, N. A. Langberg, T. H. Savits and M. Shaked, Multivariate phase-type distributions, *Opns. Res.* **32** (1984), 688–702.

5. R. Barlow and F. Proschan, *Statistical Theory of Reliability and Life Testing: Probability Models*, To Begin Again, Silver Spring, Md., 1981.

6. E. Brockmeyer, H. L. Halstrom and A. Jensen, *The Life and Works of A. K. Erlang*, Copenhagen Telephone Company, Copenhagen, 1948.

7. O. Bronshtein, I. Gertsbakh, B. Pittel and S. Shahaf, One-node closed multichannel service system: several types of customers and service rate, and random pick-up from the waiting line, *Adv. Appl. Prob.* **19** (1987), 487–504.

8. J. A. Buzacott and J. G. Shanthikumar, *Stochastic Models of Manufacturing Systems*, Prentice Hall, Englewood Cliffs, New Jersey, 1993.

9. X. Chao, Monotone effect of dependency between interarrival and service times in a simple queueing system, *Oper. Res. Lett.* **17** (1995), 47–51.

10. B. D. Choi, K. B. Choi and Y. W. Lee, M/G/1 retrial queueing systems with two types of calls and finite capacity, *Queueing Systems* **19** (1995), 215–229.

11. Q. H. Choo and B. W. Conolly, Waiting time analysis for a tandem queue with correlated service, *Eur. J. Opnl. Res.* **4** (1980), 337–345.

12. B. W. Conolly, The waiting time process for a certain correlated queue, *Opns. Res.* **16** (1968), 1006–1015.

13. B. W. Conolly and N. Hadidi, A correlated queue, *J. Appl. Prob.* **6** (1969), 122–136.

14. B. W. Conolly and Q. H. Choo, The waiting time process for a generalized correlated queue with exponential demand and service, *SIAM J. Appl. Math.* **37** (1979), 263–275.

15. D. R. Cox and W. L. Smith, On the superposition of renewal processes, *Biometrika* **41** (1954), 91–99.

16. C. F. Daganzo and E. Diez-Roux, The variance of the number of customers in an infinite-server queueing system with an arbitrarily correlated arrival process, *Queueing Systems* **19** (1995), 289–299.

17. N. M. Dijk and B. F. Lamond, Simple bounds for finite single-server exponential tandem queues, *Opns. Res.* **36** (1988), 470–477.

18. R. L. Disney and W. E. Mitchell, A solution for queues with instantaneous jockeying and other customer selection rules, *Naval Res. Logist. Quart.* **17** (1970), 315–325.

19. R. L. Disney and D. König, Queueing networks: a survey of their random processes, *SIAM Review* **27** (1985), 335–403.

20. L. W. Dowdy, B. M. Carlson and A. T. Krantz, Single-class bounds of multi-class queueing network, *J. ACM* **39** (1992), 188–213.

21. F. Downton, Bivariate exponential distributions in reliability theory, *J. Royal Statist. Soc.* **(B)** **32** (1970), 408–417.

22. E. A. Elsayed and A. Bastani, General solutions of the jockeying problem, *Eur. J. Opnl. Res.* **22** (1985), 387–396.

23. J. D. Esary and A. W. Marshall, Multivariate distributions with exponential minimums, *Ann. Statist.* **2** (1974), 84–98.

24. G. I. Falin, J. R. Artalejo and M. Martin, On the single server retrial queue with priority customers, *Queueing Systems* **14** (1993), 439–455.

25. A. Federgruen and H. Groenevelt, M/G/c queueing systems with multiple customers: characterization and control of achievable performance under nonpreemptive priority rules, *Mgmt. Sci.* **34** (1988), 1121–1138.

26. J. E. Freund, A bivariate extension of the exponential distribution, *J. Am. Stat. Assoc.* **56** (1961), 971–977.

27. F. Gillent and G. Latouche, Semi-explicit solutions for M/PH/1-like queueing systems, *Eur. J. Opnl. Res.* **13** (1983), 151–160.

28. D. Gross and C. M. Harris, *Fundamentals of Queueing Theory*, John Wiley & Sons, 1974.

29. E. J. Gumbel, Bivariate exponential distributions, *J. Am. Stat. Assoc.* **55** (1960), 698–707.

30. L. Gün and A. M. Makowski, Matrix-geometric solution for two node tandem queueing systems with phase-type servers subject to blocking and failures, *Commun. Statist. -Stochastic Models* **5** (1989), 431–457.

31. N. Hadidi, Queues with partial correlation, *SIAM J. Appl. Math.* **40** (1981), 467–475.

32. N. Hadidi, Further results on queues with partial correlation, *Opns. Res.* **33** (1985), 203–209.

33. F. A. Haight, Two queues in parallel, *Biometrika* **45** (1958), 401–410.

34. A. G. Hawkes, A bivariate exponential distribution with applications to reliability, *J. Royal Statist. Soc.* **(B)** **34** (1972), 129–131.

35. C. Heavey, H. T. Papadopoulos and J. Browne, The throughput rate of multistation unreliable production lines, *Eur. J. Opnl. Res.* **68** (1993), 69–89.

36. T. Homma and T. Fujisawa, Some notes on the queues with multiple inputs, *Yokohama Math. J.* **12** (1964), 1–15.

37. G. H. Hunt, Sequential arrays of waiting lines, *Opns. Res.* **11** (1956), 674–683.

38. E. P. C. Kao and C. Lin, A matrix-geometric solution of the jockeying problem, *Eur. J. Opnl. Res.* **44** (1990), 67–74.

39. T. Katayama, Analysis of a tandem queueing system with gate attended by a moving server, *Review of the Electrical Communi. Lab.* **29** (1981), 254–267.

40. T. Katayama, Analysis of a tandem queueing system with gate attended by a moving server with walking time, *Trans. of IECE of Japan* **J64-B, 9** (1981), 931–938.

41. T. Katayama, Analysis of a finite intermediate waiting room tandem queue attended by a moving server with walking time, *Trans. of IECE of Japan* **E-64, 9** (1981), 571–578.

42. T. Katayama, Mean sojourn times in a multi-stage tandem queue served by a single server, *J. Oper. Res. Soc. Japan* **31** (1988), 233–251.

43. T. Katayama and K. Kobayashi, Sojourn time analysis for a cyclic-service tandem queueing model with general decrementing service, *Mathl. Comput. Modelling* **22** (1995), 131–139.

44. F. P. Kelly, Networks of queues with customers of different types, *J. Appl. Prob.* **12** (1975), 542–554.

45. W. F. Kibble, A two-variate gamma type distribution , *Sankhyā* **5** (1941), 137–150.

46. L. Kleinrock, *Communication Nets; Stochastic Message Flow and Delay*, McGraw-Hill, New York, 1964.

47. E. Koenigsberg, On jockeying in queues, *Mgmt. Sci.* **12** (1966), 412–436.

48. D. König and V. Schmidt, Relationships between time/customer stationary characteristics of tandem queues attended by a single server, *J. Oper. Res. Soc. Japan* **27** (1984), 191–204.

49. T. C. T. Kotiah and N. B. Slater, On two-server Poisson queues with two types of customers, *Opns. Res.* **21** (1973), 597–603.

50. A. S. Krishnamoorthy and M. Parthasarty, A multivariate gamma type distribution, *Ann. Math. Statist.* **22** (1951), 549–557.

51. V. G. Kulkarni, A game theoretic model for two types of customers competing for service, *Oper. Res. Lett.* **2** (1983), 119–122.

52. C. Langaris, A correlated queue with infinity many servers, *J. Appl. Prob.* **23** (1986), 155–165.

53. C. Langaris, Busy-period analysis of a correlated queue with exponential demand and service, *J. Appl. Prob.* **24** (1987), 476–485.

54. G. Latouche and M. F. Neuts, Efficient algorithmic solutions to exponential tandem queues with blocking, *SIAM J. Alg. Disc. Meth.* **1** (1980), 93–106.

55. G. Latouche, On a Markovian queue with weakly correlated interarrival times, *J. Appl. Prob.* **18** (1981), 190–203.

56. G. Latouche, Perturbation analysis of a phase-type queue with weakly correlated arrivals, *Adv. Appl. Prob.* **20** (1988), 896–912.

57. S. S. Lavenberg and M. Reiser, Stationary state probabilities at arrival instants for closed queueing networks with multiple types of customers, *J. Appl. Prob.* **17** (1980), 1048–1061.

58. A. W. Marshall and I. Olkin, A multivariate exponential distribution, *J. Am. Stat. Assoc.* **62** (1967), 30–44.

59. L. E. Meester and J. G. Shanthikumar, Concavity of the throughput of tandem queueing systems with finite buffer storage space, *Adv. Appl. Prob.* **22** (1990), 764–767.

60. C. R. Mitchell, A. S. Paulson and C. A. Beswick, The effect of correlated exponential service times on single server tandem queue, *Naval Res. Logist. Quart.* **24** (1977), 95–112.

61. C. R. Mitchell and A. S. Paulson, M/M/1 queues with interdependent arrival and service processes, *Naval Res. Logist. Quart.* **26** (1979), 47–56.

62. J. J. Moder and C. R. Phillips, Jr., Queueing with fixed and variable channels, *Opns. Res.* **10** (1962), 218–231.

63. P. M. Morse, *Queues, Inventories and Maintenance*, John Wiley & Sons, 1958.

64. K. Murakami and G. Nakamura, A model for event handling in a functionally dedicated processor and analysis, *Tras. of IECE of Japan* **J61-D, 7** (1978), 465–472.

65. K. Murari, Some c-additional service channels, limited space queueing problem, *Zeit. angew. Math. Mech.* **51** (1971), 315–318.

66. S. S. Nair, Semi-Markov analysis of two queues in series attended by a single server, *Bull. Soc. Math. Belgique* **22** (1970), 355–367.

67. S. S. Nair, A single server tandem queue, *J. Appl. Prob.* **8** (1971), 95–109.

68. S. S. Nair, Two queues in series attended by a single server, *Bull. Soc. Math. Belgique* **25** (1973), 160–176.

69. R. T. Nelson, Dual-resource constrained series service systems, *Opns. Res.* **16** (1968), 324–341.

70. M. F. Neuts, The M/G/1 queue with several types of customers and change-over times, *Adv. Appl. Prob.* **9** (1977), 604–644.

71. M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*, Johns Hopkins University Press, Baltimore, 1981.

72. M. F. Neuts, Matrix-analysis methods in queueing theory, *Eur. J. Opnl. Res.* **15** (1984), 2–12.

73. M. F. Neuts, The caudal characteristic curve of queues, *Adv. Appl. Prob.* **18** (1986), 221–254.

74. T. Nishida, R. Watanabe and A. Tahara, Poisson queue with correlated two servers, *Technol. Repts. Osaka Univ.* **24** (1974), 403–409.

75. T. Nishida, A. Tahara and Y. Doi, Optimal allocation of service rate for two-stage tandem queue, *Technol. Repts. Osaka Univ.* **24** (1974), 411–416.

76. T. Nishida, T. Hiramatsu and H. Itsuji, Commutative tandem queue with Markovian input and services, *Math. Japonica* **21** (1976), 401–409.

77. T. Nishida and T. Hiramatsu, Commutative tandem queue with finite waiting room, *J. Oper. Res. Soc. Japan* **20** (1977), 194–202.

78. T. Nishida and T. Hiramatsu, Commutative tandem queue with correlated two servers, *Math. Japonica* **24** (1979), 73–80.

79. T. Nishida, Optimal allocation of service rates for interchangeable parallel two-stage tandem queue, *Math. Japonica* **41** (1995), 211–215.

80. C. A. O'Cinneide and A. E. Raftery, A continuous multivariate exponential distribution that is multivariate phase type, *Statistics & Probability Letters* **7** (1989), 323–325.

81. H. T. Papadopoulos, C. Heavey and M. E. J. O'Kelly, Throughput rate of multistation reliable production lines with interstation buffers:(I)Exponential case, *Computers in Industry* **13** (1990), 229–244.

82. H. T. Papadopoulos, C. Heavey and M. E. J. O'Kelly, Throughput rate of multistation reliable production lines with interstation buffers:(II)Erlang case, *Computers in Industry* **13** (1990), 317–335.

83. H. T. Papadopoulos and M. E. J. O'Kelly, Exact analysis of production lines with no intermediate buffers, *Eur. J. Opnl. Res.* **65** (1993), 118–137.

84. H. T. Papadopoulos, C. Heavey and J. Browne, *Queueing Theory in Manufacturing Systems: Analysis and Design*, Chapman & Hall, 1993.

85. A. S. Paulson, A characterization of the exponential distribution and a bivariate exponential distribution, *Sankhyā* **(A) 35** (1973), 69–78.

86. H. G. Perros, *Queueing Networks with Blocking*, Oxford University Press, New York, 1994.

87. M. Pinedo and R. W. Wolff, A comparison between tandem queue with dependent and independent service times, *Opns. Res.* **30** (1982), 464–479.

88. A. E. Raftery, A continuous multivariate exponential distribution, *Commun. Stat. -Theor. Meth.* **13** (1984), 947–965.

89. V. Ramaswami and G. Latouche, A general class of Markov processes with explicit matrix-geometric solutions, *OR Spektrum* **8** (1986), 209–218.

90. V. Ramaswami, Nonlinear matrix equations in applied probability-solution techniques and open problems, *SIAM Review* **30** (1988), 256–263.

91. J. Romani, Un modelo de la teoria de colas con numero variable de canales, *Trabajos estadistica* **8** (1957), 175–189.

92. M. Shaked and J. G. Shanthikumar, *Stochastic Orders and Their Applications*, Johns Academic Press, New York, 1994.

93. J. G. Shanthikumar and M. A. Jafari, Bounding the performance of tandem queues with finite buffer spaces, *Ann. Oper. Res.* **48** (1994), 185–195.

94. T. Shioyama, Optimal control of a queueing network system with two types of customers, *Eur. J. Opnl. Res.* **52** (1991), 367–372.

95. K. B. Singh, A queueing system with random additional service facility, *R. A. I. R. O.* **12** (1978), 311–318.

96. N. B. Slater and T. C. T. Kotiah, The steady state of a multi-server mixed queue, *Adv. Appl. Prob.* **5** (1973), 614–631.

97. J. H. A. Smit, The queue GI/M/s with customers of different types or the queue, *Adv. Appl. Prob.* **15** (1983), 392–419.

98. R. Szekli, R. L. Disney and S. Hur, MR/GI/1 queues with positively correlated arrival stream, *J. Appl. Prob.* **31** (1994), 497–514.

99. H. Takagi, *Analysis of Polling System*, The MIT Press, 1985.

100. M. Taube-Netto, Two queues in tandem attended by a single server, *Opns. Res.* **25** (1977), 140–147.

101. S. D. Wicksell, On correlation functions of type III, *Biometrika* **25** (1933), 121–133.

102. R. W. Wolff, Tandem queues with dependent service times in light traffic, *Opns. Res.* **30** (1982), 619–635.

103. C. H. Yoon and C. K. Un, Performance of personal portable radio telephone system with and without guard channels, *IEEE J. Select. Areas Commun.* **11** (1993), 911–917.

List of Publications

1. T. Nishida and K.Yoneyama, Correlated multiserver queue with random input, *Math. Japonica* **22** (1977), 395–401.

2. T. Nishida, Y. Ueshima and K. Yoneyama, Two-stage tandem queue with two types of customers, *Technol. Repts. Osaka Univ.* **28** (1978), 343–349.

3. T. Nishida, Y. Serikawa and K. Yoneyama, n queues in tandem attended by a single server, *Rep. Stat. Appl. Res., JUSE* **26** (1979), 14–19.

4. J. M. Huang, T. Nishida and K. Yoneyama, Commutative tandem queue with two types of customers, *Math. Japonica* **25** (1980), 533–543.

5. K. Yoneyama, T. Yuasa and T. Nishida, A queueing system with c-random additional service channels, *Zeit. angew. Math. Mech.* **60** (1980), 57–59.

6. K. Yoneyama, Two stage tandem queue with correlated service times, *Res. Repts. Himeji College of Hyogo* **40** (1995), 37–40.

7. K. Yoneyama and T. Nishida, A note on the concavity or the convexity of performance measures of two-stage finite tandem queue, *Math. Japonica* **42** (1995), 213–218.

8. K. Yoneyama and H. Ishii, The effect of correlated exponential service times on queues, *Technol. Repts. Osaka Univ.* **45** (1995), 101–106.

9. K. Yoneyama, The throughput rate of commutative tandem queue, *Res. Repts. Himeji College of Hyogo* **41** (1996), 135–140.

10. K. Yoneyama and H. Ishii, Correlated multiserver queue with instantaneous jockeying, *Technol. Repts. Osaka Univ.* **46** (1996), 1–7.

11. K. Yoneyama, The waiting time distribution of correlated multiserver queue, *Math. Japonica* **43** (1996), 569–576.

12. K. Yoneyama and H. Ishii, The throughput rate of interchangeable parallel two-stage tandem queue with correlated service times, *Computers Math. Applic.* (to appear).