ANALYTICAL MODELS FOR PERFORMACE EVALUATION OF PRODUCTION-INVENTORY SYSTEMS

By

SHANKAR SIVARAMAKRISHNAN

Master of Management Studies
Birla Institute of Technology and Science
Pilani, INDIA
July, 1989

Master of Science
Oklahoma State University
Stillwater, Oklahoma
May, 1993

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements of
the Degree of
DOCTOR OF PHILOSOPHY
July, 1998

ANALYTICAL MODELS FOR PERFORMACE

EVALUATION OF PRODUCTION-

INVENTORY SYSTEMS

Thesis Approved:

Manjunath Kamath

Thesis Advisor

David B. Pratt

Kenneth E Case

Michael Branson

R. Ramakumar.

Wayne B. Powell

Dean of Graduate College

ACKNOWLEDGMENTS

I wish to dedicate this work to the following teachers who have played a great role in helping me gain wisdom and molding my life.

First and foremost I dedicate this work to God, the Almighty, who in the form of Dakshina Kalika has provided me with the strong mind to complete this journey. I would like to lay this work at HER feet in gratitude for everything that SHE has bestowed upon me. Next, I dedicate this work to my guru, Dr. Manjunath Kamath, who not only has been my teacher but also a friend and mentor who provided me with valuable guidance and emotional support. I would like then dedicate this work to my dearest wife, Hema Lakshmanan, who not only helped me in many aspects of my dissertation work, and just being there for me all the time during the course of this journey. Lastly, I dedicate this work to my parents Mr. Sivaramakrishnan and Mrs. Vijayalakshmi Sivaraman, and my brother, Mr. SankaraRaman, without whose efforts and support, this journey would not have been possible.

Special thanks go to all the members of my doctoral committee for the way they have contributed in building and impacting my thought process towards my research. Dr. Ken Case has been responsible during my early part of program to build a solid foundation in quality control, reliability and statistics. Dr. Pratt has provided numerous suggestions in the understanding of the simulation aspects. Dr. Branson has been instrumental in teaching me the fundamentals of stochastic process and statistical aspects of simulation. Dr. Ramakumar has been instrumental in my understanding of several key reliability concepts from the research conducted by electrical engineers.

Among friends and collegues, I would like to express my heart felt gratitude to Baskar Krishnamoorthy, Partha Ramachandran, and Ananth Krishnamoorthy for the tremendous help on not only the technical aspects but also for helping me in several other things during the long distance work from Houston.

Finally, I would like to acknowledge the financial support provided by National Science Foundation during the last four years of my study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GLOSSARY OF TERMS

The definitions of a selected list of technical terms are included in this glossary to clarify their intended meaning and usage.

Homogeneous system	A system which has the same basestock level and the same processing time distribution at all the stages.
Intermediate parts	Parts that are stored in the stores on the output side of the stages ahead of demand. These are parts which have been processed at one or more stages. These are different from purchased parts that are directly used in intermediate processing steps in some manufacturing systems.
SCV	The squared coefficient of variation which is defined as the ratio of the variance of a random variable to the square of its mean.
WIP	Expanded as work-in-process, WIP is defined as the orders with or without parts that await to be processed at the input side of a stage.
Basestock level	This policy parameter determines the maximum planned inventory of an item at a stage.

CHAPTER I

INTRODUCTION

Manufacturing systems have existed for a long time. A pragmatic purpose of a manufacturing system is to meet customer demand for high quality and reliable products at minimum cost. Production of goods involves interaction among several components, both external (e.g. suppliers) and internal (e.g. machines and material handlers) to the manufacturing system which have to work with each other to achieve this common goal. Competitiveness in the global market is forcing manufacturing organizations to be very flexible with respect to demand and product mix changes to just stay in business. In such a dynamic environment, modeling of the underlying production systems becomes critical for their effective design and control.

During the lifetime of any manufacturing system, the firm responsible for it goes through many phases of decision making, from an analysis of initial feasibility through detailed design of the facility; installation and startup; and final obsolescence of the facility (Suri, 1988). Many decisions have to be made on a routine basis in any manufacturing facility such as those related to capacity, amount of tooling and fixtures, and amount and location of storage spaces. Performance evaluation techniques help the decision-maker in making these key decisions during the design phase as well as the operational phase of these systems (Viswanadham and Narahari, 1992). Also, these techniques provide an insight into the behavior of the manufacturing system and help us gain a better understanding of the dynamics of the system.

In this research, we consider discrete part manufacturing systems; items are produced as discrete units in these systems. Such systems are commonly found in mechanical, electrical, and electronics industries.

1.1 PERFORMANCE EVALUATION OF MANUFACTURING SYSTEMS

Performance evaluation involves the development and solution of models for determining the values of the performance measures that can be expected from a given set of decisions (Suri et al., 1993). During the last several years, many researchers have contributed towards the understanding of complexities present in manufacturing systems through the use of a variety of modeling tools such as simulation, Markov chains, Petri nets, and queueing. To a great extent, the previous efforts focused on a limited set of issues within a single model. *To support decision making in today's dynamic environment, a larger set of issues needs to be included within a single model so that the impact of their interactions on the total system performance can be evaluated.* Hence, performance evaluation of manufacturing systems continues to remain a challenging and active research area.

There are several schemes for classifying performance evaluation models. A commonly used scheme in the context of discrete part manufacturing systems is to classify them as simulation and analytical models. Simulation models represent the events that occur as a system evolves by a sequence of steps in a computer program. The probabilistic nature of the events is modeled by sampling from distributions representing the timing and pattern of occurrence of such events. Analytical models describe the system using mathematical or symbolic relationships (Buzacott and Shanthikumar, 1993). These analytical models are also called aggregate dynamical models (Suri et al., 1993) since they capture the stochastic nature of the systems in an aggregate manner.

Analytical models are increasingly being used in industry for rapid decision making purposes (Segal and Whitt, 1989; Suri and de Treville, 1993). Their main advantages compared to simulation are modeling ease and speed. However, many assumptions are required to obtain tractable analytical models. Hence, these models are appropriate for rapid and rough cut analysis (Kamath, 1994). Many publications have appeared that suggest ways of exploiting the complementary nature of the analytical and simulation approaches. The basic idea is to combine the speed of analytical models with the detailed modeling capability of simulation. For example, analytical models can be used to quickly eliminate a large number of design alternatives to provide a handful of potential ones which can then be investigated in detail by simulation (Suri and Diehl, 1987; Suri and de Treville, 1991).

The development of analytical models for manufacturing began in the late fifties with the seminal works of Jackson (1957) for job shops and Koenigsberg (1959) for cyclic systems. These papers were followed by numerous research publications that expanded this area. A comprehensive review of the developments in the analytical modeling area is contained in Suri and de Treville (1993). Recently, several textbooks have been published that focus on manufacturing systems modeling and analytical modeling in particular (Askin and Standridge, 1993; Buzacott and Shanthikumar, 1993; Gershwin, 1994; Viswanadham and Narahari, 1992).

1.2 MOTIVATION BEHIND THIS RESEARCH

The application of queueing theory in industry is not as widespread as it should be because many of the early models contained assumptions which were viewed as too restrictive by many industries. Research in the last decade has been largely devoted to obtaining approximate solutions to more exact models, and as a result, there has been a resurgence of research in queueing applications in manufacturing systems. Also, queueing models seem to be gaining a wider acceptance in industry mainly due to the accessibility to such models via software packages like MPX (Suri et al., 1995). Typically, in a queueing network model, customers/parts visit several nodes/workstations before they depart the system. These networks model manufacturing systems where parts are *made-to-order,* and an order typically visits various nodes/workstations as required by the sequence of operations for that order. These types of models are well suited for dealing with *capacity* and *congestion* issues. However, in many manufacturing systems, short-term capacity issues are often tackled by holding an inventory of finished goods and intermediate parts at the output side of workstations to counter the demand. Queueing models do not consider such planned inventories and usually assume an unlimited supply of raw materials and zero intermediate parts.

Inventory theory has been studied since 1913 with the development of the famous EOQ model by F.W. Harris. Ever since, a vast array of models has been developed that include many possible complexities and less restrictive assumptions compared to the earlier models; numerous books have been published on this subject. Essentially, the main objectives of these models have been to determine optimal ordering quantities as well as holding inventories. Traditionally, inventory models have ignored capacity and congestion issues, i.e., these models assume that the production system has infinite capacity.

Recently, Lee and Zipkin (1992), Zipkin (1995a, 1995b) and Buzacott and Shanthikumar (1993) have developed models that include congestion and capacity issues as well as planned inventories of both intermediate parts and finished goods. These models are suitable for *make-to-stock* systems which are quite prevalent in many discrete part manufacturing industries. Several unanswered questions remain with respect to the modeling of make-to-stock systems which makes it an active research area.

The aforementioned models of production-inventory systems fit into the broader framework of *supply chain* models. A supply chain is a network of facilities that performs the functions of procurement of material, transformation of material to intermediate and finished products, and distribution of finished products to customers (Lee and Billington, 1993). There have been many significant developments in the supply chain management area, and they have contributed to the success of many companies (Lee and Billington, 1995). However, many of the supply chain models that have been developed do not include the congestion effects due to limited capacity. In fact, Lee and Billington (1993) mention this as a potential research issue that requires immediate attention. Models that can simultaneously handle queueing and inventory issues would be extremely useful in designing and managing a supply chain.

1.3 THE PROBLEM STATEMENT

It is evident that the analysis of production-inventory systems is very crucial to the success of several manufacturing systems. Its importance to supply chain has been clearly shown in the earlier section. Though there is abundant literature on production-inventory systems, the existing models do not adequately address the issues of inventory and capacity/congestion within a single modeling framework. This research has addressed these issues within a single framework along with some reliability issues. The problem statement can be described as "developing analytical models for production-inventory systems that simultaneously address inventory, congestion/capacity and reliability issues."

1.4 THE PROPOSED RESEARCH AREA

The main focus of this research is the study of production systems with planned inventories. The approach developed builds on the parametric decomposition approach that has proven to be quite successful in dealing with queueing models of manufacturing systems. This approach is described in detail in Chapter II. Lee and Zipkin (1992) in their study of the tandem queues with planned inventories, used an approach developed by Svoronos and Zipkin (1991) for the analysis of multi-echelon inventory systems. The multi-echelon models impose many restrictions on Lee and Zipkin's overall approach and so far, they have examined only a restricted class of systems. Our approach seems to be more robust and the systems of Lee and Zipkin (1992) become a subset of the wider range of systems that can be modeled by our approach.

The set of performance measures which we calculate is similar to that used by Lee and Zipkin (1992), Zipkin (1995b) and Buzacott and Shanthikumar (1993). These measures include the expected number of backorders at each stage, the expected inventory level at each stage and the expected intermediate inventory in the system.

1.5 OVERVIEW OF THE DISSERTATION

The remainder of this dissertation is presented in six chapters. Chapter 2 reviews the various research efforts that focus on the use of queueing models in modeling production/inventory systems. This chapter also summarizes some key research contributions related to the parametric decomposition approach, which is used in the analysis of queueing network models. This chapter also includes a brief introduction to the area of performability analysis, a combined analysis of performance and reliability and some relevant literature on the use of this approach in manufacturing systems. Research goals, research objectives and the research plan are outlined in Chapter 3 along with the scope and limitations of this study. Chapter 4 presents the proposed decomposition approach in detail for a tandem configuration which forms the basis for the analysis of more complex systems. This procedure is then extended to include general arrivals and general service times which is outlined in Chapter 5. Chapter 6 takes the framework developed in the previous chapter and applies it to model systems that include other manufacturing features, such as multiple servers, batch service, multiple part types, and failures of machines. Detailed numerical investigations are also presented in this chapter. Chapter 7 discusses the method for analyzing feed-forward systems and tandem systems with feedback. Chapter 8, the concluding chapter, summarizes the main research contributions along with some directions for future research.

CHAPTER II

LITERATURE REVIEW

2.1 CHAPTER OVERVIEW

In this chapter, we present a review of the literature on the analysis of production-inventory systems and the approximation methods available for analyzing queueing networks. The contributions that served as the foundation for the current research are those by Lee and Zipkin (1992), Buzacott and Shanthikumar (1993) and Whitt (1983). Section 2.2 briefly discusses the use of queueing models in analyzing production-inventory systems. The next section details the methodology of Lee and Zipkin (1992) used in the evaluation of tandem make-to-stock systems with Poisson arrivals and exponential processing times. Section 2.4 discusses the various single-stage production-inventory systems analyzed by Buzacott and Shanthikumar (1993). It is then followed by a review of the parametric decomposition approach for analyzing queueing networks (Whitt, 1983). The concluding section discusses the concept of performability and some relevant literature in that area.

2.2 QUEUEING MODELS IN PRODUCTION-INVENTORY SYSTEMS

Queueing models are well suited for studying make-to-order systems. In queueing networks, orders/customers visit nodes/workstations before they depart the system. In these type of networks, the congestion measures model the waiting before service/processing at a node/workstation. Modeling the availability of raw materials and intermediate parts required for processing is not an issue, because the models assume an unlimited supply of raw materials and no intermediate parts. On the other hand, inventory models do not usually model the effects of congestion and assume capacity to be unlimited.

The recognition that production-inventory systems can be modeled as queueing systems is attributed to Morse (1958). He treated the production system as equivalent to an infinite number of parallel servers. He used an $M/G/\infty$ model if backorders were permitted. With lost sales he used an M/G/Z/Z model, where Z is the maximum stock at the store. Sherbrooke (1968) used a similar approach called the METRIC approach in which the production-inventory system is modeled as an $M/D/\infty$ system, and exact expressions for backorders and inventory distributions are obtained. Queueing results have also been used in production-distribution systems and multi-echelon inventory systems (Federgruen, 1993; Muckstadt and Roundy, 1993). Zipkin (1984) used a combination of standard inventory models and queueing sub-models to determine the batch sizes and safety stocks in a multi-item batch production system. Despite the enormous literature on production-inventory systems (Altiok 1989; Altiok and Ranjan, 1995; Altiok and Shiue, 1994; Gavish and Graves, 1980; Goyal and Gunasekaran 1990), queueing models have seldom been used for performance evaluation that focuses on modeling congestion due to limited capacity in such systems (Buzacott and Shanthikumar, 1993).

So far, the major contributions to the performance evaluation of production-inventory systems using queueing theory are due to Zipkin, Buzacott, and Shanthikumar. Lee and Zipkin (1992) used queueing results to develop an approximation for the performance evaluation of a tandem line with planned inventories, exponential processing times, and Poisson demand. Zipkin (1995b) extended this work to tandem queues with feedback and planned inventories. Zheng and Zipkin (1990) and Zipkin (1995a) developed a queueing model to analyze the value of centralized inventory information. Buzacott and Shanthikumar (1993) present several exact and approximate models for a variety of single-stage, make-to-stock systems. The next few sections summarize these contributions in some detail, because of their importance to the research carried out in this dissertation effort.

2.3 LEE AND ZIPKIN'S MODEL FOR TANDEM QUEUES WITH PLANNED INVENTORIES

The focus of Lee and Zipkin (1992) was on tandem make-to-stock systems. Arriving customers demand a final product, and the demand is satisfied from the finished goods inventory, if available. Lee and Zipkin (1992) focused on the special case where the customer demand process is Poisson; the replenishment policy is one-for-one; and the processing times are mutually independent and exponentially distributed with the same distribution at each stage. They assumed that the system is controlled by a stationary *demand-pull* or *basestock policy*. A policy of this kind is specified by the non-negative, integer parameters S_j, $j = 1, 2, ..., J$. The quantity S_j is called the basestock level for stage j, and it determines the maximum planned inventory at the output side of stage j. S_J denotes the maximum finished goods inventory. Each stage can be thought of as operating its own local production-inventory control system; a customer demand is viewed as occurring at stage J, and a demand at each stage immediately triggers a demand at its predecessor; thus, each customer demand creates a demand at every stage. Stage J (the last stage) fills the customer demand if there is a finished unit available; otherwise the demand is backordered. The same process occurs at each stage; if stock is available, the demand is filled, else a backorder is logged.

The customer demand, if fulfilled, triggers an immediate order to replenish the inventory. This order then looks for a part in the previous stage's output and if available, goes and waits for processing. If a part is not available, the order waits for a part to arrive from the previous stage. The queue at the processing stage is assumed to be infinite. Units after completing processing move to the output buffer or to the subsequent queue in response to the demands at the present stage. If there are outstanding backorders at a stage when it completes processing of a unit, that unit is immediately released to fulfill one of the backorders. In effect, each stage j in the tandem line first works down its backorder log and then works to fill its output buffer to the basestock level S_j.

If all S_j's are zero, the system operates just like an ordinary make-to-order tandem line. This special case can be solved exactly using Jackson's (1957) product-form result. Hybrids of make-to-stock and make-to-order systems, where customer specific features are added to units at some intermediate stages, can be represented by constraining certain of the S_j's to be zero. Traditional models of multi-echelon systems do not explicitly consider limited production capacities and congestion. The properties of Jackson (1957) network are violated whenever any of the S_j, $j <$ J are greater than zero, and it becomes difficult to obtain an exact solution. Lee and Zipkin (1992) capture the congestion measures from the queueing model and then use an approximation scheme developed by Svoronos and Zipkin (1991).

2.3.1 The Approximation of Svoronos and Zipkin

The approximation used by Lee and Zipkin employs one of the multi-echelon models developed by Svoronos and Zipkin (1991). It is thus important to understand the technique of Svoronos and Zipkin for modeling a multi-echelon inventory system.

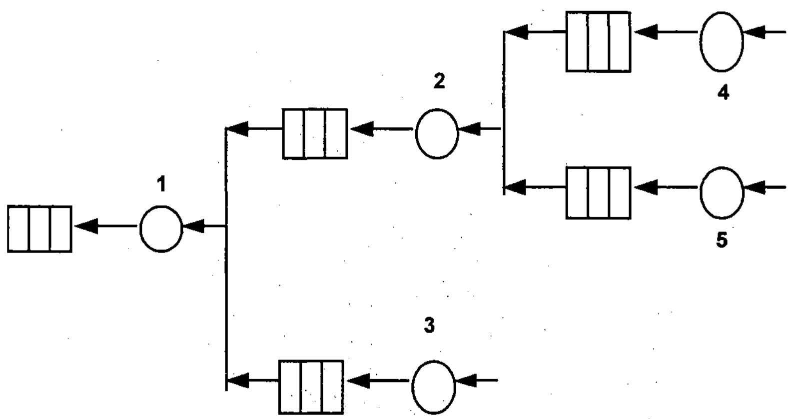


Figure 2. 1: A Multi-Echelon Inventory System

A multi-echelon inventory system consists of several facilities or locations whose supply-demand relationships form a hierarchy. There is a single location at the highest level of the hierarchy, called the central depot, whose orders go to an outside store. The lowest level of the hierarchy are the leaves of the tree, where exogenous demands occur. Figure 2.1 illustrates a multi-echelon system. Stage 1 supplies stage 2 and stage 3. Stage 2 in turn supplies stages 4 and 5. The demands occur at the stages 3, 4, and 5. Demands consume products at the leaves, and in turn the leaf stages place a demand at the predecessor stages which will fulfill this demand if they have inventory. A predecessor stage in turn will trigger a demand to its predecessor stage and this process continues till the central depot. The central depot in turn orders from the outside source and it is assumed that the outside source has ample stock. The time to fulfill an order without any delay to its successor is called the transit time. These transit times can be production times or just transportation times. Svoronos and Zipkin (1991) extend their approach of a single location model to multi-echelon systems. The following procedure is used to analyze the single location model.

Let I denote the inventory at the stage; B - the number of backorders; K - the number of outstanding orders which is the sum of the replenishment orders and backorders at the stage; and S the basestock level at the stage.

$$K = S - I + B \tag{2.1}$$

Let T denote the transit time at the stage; D represent the delay which is the time it takes to obtain an order from the outside source (or predecessor stage in multi-level systems); and L denote the total lead time at the stage. Let F_T, F_D and F_L denote the distributions of the transit time, the delay and the lead time, respectively. Let the demand at this single stage location be a Poisson process with rate λ. A one-for-one replenishment policy with a basestock level S, is assumed; i.e., every order consumed generates a demand for replenishment.

The steady state behavior is characterized by the densities of I, B and K. The other performance measure of interest is the time delay for a backorder. Some of the key results from the Svoronos and Zipkin (1991) paper are presented next.

The variable K has the same distribution as the lead-time demand, the number of demands in a random time with distribution F_L. The variable B has the same distribution as the customer-delay-demand, the distribution denoted by F_D.

This single location procedure is applied recursively, starting at the root stage (the highest echelon) and working down, to analyze the entire system. At the root stage L = T and thus F_L is known. At any other stage $j \in$ Successor (k) $F_{L_j} = F_{D_k} * F_{T_j}$ where * denotes convolution. For example, in Figure 2.1, the distribution of K at stage 2 is $F_{L_2} = F_{D_1} * F_{T_2}$. If all the transit times follow phase-type distribution and the demand is Poisson, the lead-time distribution which is a convolution of two phase-type distributions is also a phase-type distribution.

2.3.2 Application of Svoronos and Zipkin's method in the analysis of tandem queues with planned inventories.

The tandem system is a special case of the multi-echelon system. Stage j is first treated as an *M/M/1* queue in isolation. The average sojourn time is $1/(\mu_j - \lambda)$ where μ_j is the processing rate at stage j and λ is the demand rate. After determining the average sojourn times at each of the stages individually, the approach of Svoronos and Zipkin (1991) is applied to obtain the distribution of backorders and inventories at each stage. As the sojourn times are exponentially distributed they become the transit times in the multi-echelon inventory model. Lee and Zipkin (1995) used a similar idea for the analysis of sequential refinement systems. Zipkin (1995b) extended this procedure to model tandem queues with feedback. Each production stage occasionally produces a defective unit, which must then repeat processing, return to an earlier stage, or be discarded.

2.4 SINGLE-STAGE MAKE-TO-STOCK SYSTEMS

Buzacott and Shanthikumar (1993) have developed a class of models to capture the various aspects of a single-stage make-to-stock system. There are several aspects that need careful consideration in a make-to-stock system. They can be broadly classified as those pertaining to customer demand and those to the manufacturing process. Some of the critical aspects as perceived by Buzacott and Shanthikumar (1993) are presented below.

Production Variety: When several different products are made using common facilities, it is important to determine the interrelationships between the demand and manufacture of the products. In other words, it is necessary to know whether different products should be produced simultaneously or one at a time, and whether they will be demanded together or independently.

Pattern of Demand: Two aspects of demand pattern that are significant are the arrival of customers and the demand for items by a customer. The authors restrict the analysis to stationary demand patterns. But, customers may require just one item or the number of items demanded may be a random variable.

Manufacturing Capability: All manufacturing processes are to some extent unreliable or uncertain. Examples of unreliability include failure of machines, tool breakage and operator absenteeism. Also included in these systems are the quality aspects of the products. Buzacott and Shanthikumar developed models to include these features.

Buzacott and Shanthikumar restricted their analysis to single-stage systems, and therefore assumed ample supply of raw materials, parts, and tools. They developed the concept of *production authorization* cards. The authors assumed that when each item is produced by the manufacturing facility, a tag is associated with the item and thus for every item in the output store there is a tag. When a unit of a certain product is given to a customer, the tag is removed, and this then becomes the production authorization or PA card for that product. The PA card generation mechanisms together with maximum inventory level of finished goods result in a wide variety of models. The following discussion gives the procedure for a single machine with unit demand and backlogging and results are presented for the "M/M/1" model.

Single machine with unit demand and backlogging: It is assumed that there is only one machine to process the items and that customer demands that are not satisfied immediately are backlogged. It is also assumed that there are S tags available in the system. Suppose the output store is initially full at time zero and an unlimited amount of raw materials is available. Let $I(t)$ be the finished goods inventory, $B(t)$ be the number of customers backlogged, and $C(t)$ the number of PA cards available at the machine, all at time t. Let $N(t)$ be the number of jobs in the system. We have

$$I(t) = Min\{0, S - N(t)\} \tag{2.2}$$

$$B(t) = Min\{0, N(t) - S\} \tag{2.3}$$

and

$$C(t) = Min\{N(t), S\} \tag{2.4}$$

The above notations are similar to those used by Lee and Zipkin (1992) which were described in Section 2.3. It can be easily shown that

$$B(t) + C(t) = N(t) \tag{2.5}$$

Therefore, the study of the process $N(t)$ is sufficient and the information about the process $I(t)$, $B(t)$ and $C(t)$ can be derived from $N(t)$. As an example of this analogy, the results for the M/M/1 model are given below.

Results for an M/M/1 model: The production system is modeled as an M/M/1 queue where the customer arrival process is Poisson with rate λ and the processing times are exponentially distributed with mean $1/\mu$. Using the steady-state result, $P\{N = n\} = (1-\rho)\rho^n$, n=0, 1, ..., for $\rho < 1$ we get from equations 2.2 to 2.5

$$P\{B = n\} = \begin{cases} 1-\rho^{S+1}, n = 0 \\ (1-\rho)\rho^{n+S}, n = 1, 2, ... \end{cases} \tag{2.6}$$

$$P\{I = n\} = \begin{cases} \rho^{S}, n = 0 \\ (1-\rho)\rho^{S-n}, n = 1, 2, ..., S \end{cases} \tag{2.7}$$

Also,

$$E[B] = \frac{\rho^{S+1}}{1-\rho};\ E[I] = S - \frac{\rho}{1-\rho}(1-\rho^{S}) \tag{2.8}$$

Also, the steady-state probability that a customer is backlogged is $P\{\text{a customer is backlogged}\} = P\{I = 0\} = \rho^S$. This is a consequence of the PASTA property (Poisson Arrivals See Time Averages (Wolff, 1982)).

Buzacott and Shanthikumar (1993) also presented approximate results for a generalized version ("GI/G/1") of the above system. Some of the other models that have been presented by Buzacott and Shanthikumar (1993) are given below.

Single machine with unit demand and lost sales: The various cases discussed are the M/M/1/Z, M/G/1/Z, and GI/G/1/Z models.

Single machine with interruptible demand: In this case, the arrival generation process is switched off and no more arrivals can be generated as long as the output store is empty. The authors developed GI/M/1/Z and GI/G/1/Z stopped arrival models to analyze these types of systems.

Single or multiple machines with bulk demand: The above ideas were extended to a general single-stage manufacturing system where each customer requires more than one part. The authors presented an approximation for the $GI^X/G/c$ model of the above system.

Produce-to-stock with yield losses: Two scenarios have been studied in this type of system: (i) defects are detected at the manufacturing facility and the items are reprocessed until they are defect free; and (ii) defects are detected at delivery and the item is discarded.

Some of these models have been used in modeling Kanban and MRP type systems (Buzacott 1989; Buzacott and Shanthikumar, 1992).

2.5 PARAMETRIC DECOMPOSITION APPROACH FOR ANALYZING QUEUEING NETWORKS

This section presents a brief description of the parametric decomposition approach that forms the basis for the analysis approach developed in this dissertation.

In the mid eighties, a fundamental shift occurred when many researchers working in the queueing area started to focus more on the application side than on the exactness of the solution methodology. Whitt (1983) described the change from a modeling viewpoint by stating “a natural alternative to an exact analysis of an approximate model is an approximate analysis of a more exact model.” A comprehensive methodology for analyzing open queueing network models that explicitly considers the variability of both the arrival and service processes emerged. Seminal work in this area is credited to Kuehn (1979) and Whitt (1983). Whitt (1983) may be viewed as the main archival reference for details of what is now known as the *parametric decomposition* (PD) approach. Several features in addition to general arrival and service times have been modeled and the PD approach has been the basis for modeling these features. In the 1983 paper, Whitt presented the details of the PD approach in the context of a software package called the queueing network analyzer (QNA). The following description of the PD approach for an open single-class network with single-server FCFS nodes is adapted from Kamath (1994).

In an open queueing network, customers enter the network from the outside, receive service at one or more nodes and eventually leave the network. For node *i* in the network, the following variability parameters are used.

c_{0i}^2: inter-arrival time SCV of external arrivals to node *i*;

c_{si}^2: service time SCV at node *i*;

c_{ai}^2: inter-arrival SCV of total arrivals at node *i*; and

c_{di}^2: inter-departure time SCV at node *i*.

The squared coefficient of variation (SCV) of a random variable (rv) is defined as the variance of the rv divided by the square of its mean.

The PD approach involves two main steps. The first step is the analysis of the interaction between nodes to approximately determine the mean and the SCV of the inter-arrival time at each node. The next step computes the performance measures based on GI/G/*m* approximations that are based on the first two moments of the inter-arrival and service times.

Computing parameters of arrival processes at the nodes: A single node is related to other nodes in the network model by its input and output processes. The internal flow parameters (rates and variability parameters of arrival processes) approximately capture the interdependence among the nodes. First, the mean total arrival/departure rate of customers at node *i* is obtained via the traffic rate equations representing the conservation of flow. If utilization $\rho_i \geq 1$, then the i^{th} node is unstable and the procedure stops. The calculations involving the arrival rates and utilization are exact. Approximations are used while setting up of the traffic variability equations which yield the variability parameters for the internal flows, c_{ai}^2. The equations are linear, and are obtained by combining renewal approximations for the basic network operations, namely, merging of flow, splitting of flow and flow through a node. The details of these approximations can be found in Bitran and Dasu (1992), Tirupati (1992) and Whitt (1983). In summary, this step involves the solution of two sets of linear equations - the *traffic rate equations* yield the total arrival rate at each node and the *traffic variability equations* yield the SCV of inter-arrival times at each node.

Calculation of node performance measures: This procedure involves approximations developed in the queueing literature for GI/G/1 or GI/G/m queues (Kraemer and Langenbach-Belz, 1976; Shanthikumar and Buzacott, 1980; Whitt 1993). The queues at the nodes are treated as being stochastically independent and the expected waiting time at each queue is computed using approximate formulae that are based on the first two moments of the inter-arrival and service times. Using Little's law (Little 1961) other measures such as the mean queue length at the nodes can be obtained.

The PD approach is closely followed in the development of our proposed approach for analyzing production-inventory systems. The details of our approach are presented in the subsequent chapters. As mentioned earlier, this research also addresses the modeling of reliability issues in production-inventory systems. In this regard, we end this chapter with a brief review of performability analysis which combines reliability modeling and system performance evaluation.

2.6 PERFORMABILITY ANALYSIS

Traditionally, equipment availability issues have been handled by reliability theory. Performance models sometimes incorporate the delays due to minor machine disruptions into the service time and obtain approximate values for performance measures. A combined study of performance and reliability called performability modeling is applicable to the study of "fault-tolerant" systems. Structural changes could be because of a variety of reasons. Examples are machines becoming inactive because of failures, change in suppliers, and changes in labor force. The overall system is still functional though these system changes affect the performance of the system and the system can be called tolerant to such system changes.

The following definitions were adapted from Viswanadham and Narahari (1992). A *fault tolerant system* is one that has an inherent capability to adapt automatically, in a well-defined manner, to failures of its components, so as to maintain continuously a specified level of performance. Given a fault-tolerant system, a *structure state* of the system is a vector whose components describe the condition of individual subsystems as influenced by reconfigurations.

Let $Z(u)$ be a structure state of a fault tolerant system at $u \geq 0$. Then the family of random variables $\{Z(u) : u \geq 0\}$ is called the *structure state process* (SSP) of the system. Given a structure state i, its associated reward f_i is a random variable that describes the performance of the system in that structure state.

Given (i) a system with structure state process $\{Z(u) : u \geq 0\}$ having state space $S=\{0, 1, 2, ..., m\}$ and (ii) rewards $f_0, f_1, f_2, ..., f_m$ in the individual structure states, the performability $Y_t(s)$ over an observation period $[0, t]$ and with initial structure state as $s \in S$ is a random variable given by

$$Y_t(s) = \sum_{i=0}^{m} f_i \tau_i \qquad (2.9)$$

where τ_i is the total time $[0, t]$ that the SSP stays in state i. In a performability context, three measures are often computed: performability distribution, steady-state performability, and interval performability.

The performability distribution is the cumulative distribution of performability $Y_t(s)$, i.e. $P\{Y_t(s) \leq x\}$ for $x \in R$. The limit as $t \to \infty$, if it exists, is called the *steady-state performability*; and the expected value $E[Y_t(s)]$ is called the *interval performability.*

Performability analysis has been studied in the context of fault tolerant systems such as computer processors. Meyer (1980) coined the word "Performability" to signify the combined study of performance and reliability issues under the same framework. Recently, performability analysis has been used in the study of automated manufacturing systems (AMSs). Viswanadham et al. (1991) were one of the early researchers who applied this framework in manufacturing systems. Viswanadham et al. (1995) applied this framework in the study of AMS with multiple part types. Ram and Viswandham (1995) obtained the performability measures of an AMS with a centralized material handling system.

Solution procedures for finding the performability distribution for any system do not have a generalized procedure. It is quite possible that an SSP for a given system is unique, and hence requires a new solution procedure to determine the performability measures for that system. The published applications of performability in manufacturing systems have typically used existing solution procedures available for known SSPs in the literature. Viswanadham et al. (1991) use the procedure by Dontiello and Iyer (1987) for finding the performability measures of a flexible manufacturing cell with multiple machines and a centralized material handler. Pattipati (1993) used the techniques of stochastic differential equations to obtain the performability density and distribution for a non-homogeneous Markov process. Iyer et al. (1986) presented a computational method for determining moments of performability for repairable systems. Smith et al. (1988) developed an algorithm for the numerical evaluation of performability distributions in repairable systems. Finding solution techniques for the various types of the Markov and semi-Markov reward models is in itself a vast and active research area.

An example illustrating the application of this technique to manufacturing systems is now described. Consider a manufacturing system with two types of machines. One of the machines is an automatic machine and the other is a semi-automatic type. Orders are usually processed on the automatic machine. The automatic machine processes orders at twice the rate as the semi-automatic machine but is prone to failures. When the automatic machine is down for repairs, the semi-automatic machine is used to process the orders. The breakdowns are not very frequent and the repair process takes a sufficiently long time. In other words, it is reasonable to assume that processing on the semi-automatic machine reaches steady state before the repair is completed. Also, processing of orders on the automatic machine reaches steady-state between breakdowns. Thus the SSP for this system can be described to exist in two states, each state indicating the specific machine in use. Let the production rate be the performance measure of interest. A performability analysis would help answer the following types of questions.

What is the probability of producing 8,000 parts in a 3-month period?

How long is it going to take to deliver 5,000 parts with a probability of 0.90?

As described in the beginning of the section, the performability concept is applicable to production-inventory systems as well. This concludes the review of the relevant literature. The next chapter presents the research objectives of this dissertation effort.

CHAPTER III

RESEARCH OBJECTIVES

3.1 STATEMENT OF RESEARCH

Research Goal

The overall goal of this research was to develop analytical models for the performance analysis of production-inventory systems that can simultaneously address inventory, capacity/congestion and reliability issues.

The objectives that are described in the following section mostly address the development of analytical models for production-inventory systems. The dissertation discusses the development of these models in detail and examines the accuracy of the models by comparing analytical results with simulation estimates for several example systems. The next few chapters document the development of the analytical models and algorithms. They also include a summary of the numerical investigations conducted to test the accuracy of the analytical models.

Most of the objectives focus on the tandem or flow line configuration which is a very common configuration in many manufacturing systems. Also, it is typical for any new research in analytical modeling of manufacturing systems to start with the tandem configuration, and later extend the models to other configurations.

3.2 RESEARCH OBJECTIVES

Objectives pertaining to the tandem configuration.

In Objectives 1 through 6, the model development focuses on variations of the tandem configuration of make-to-stock systems. Collectively the objectives address the development of models that can handle several complexities of production-inventory systems.

OBJECTIVE 1: The objective was to develop an analytical solution methodology based on a new decomposition approach for tandem make-to-stock systems with single, reliable servers at each stage; to conduct extensive numerical investigations to test the accuracy of the analytical solutions; and to develop a general analysis framework for make-to-stock systems.

A wide variety of systems were tested by changing the parameters of the tandem configurations. Tandem configurations consisting of three and ten single-server stages with general demand arrivals and general processing times, comprising of both homogeneous (same service distribution and basestock level at all stages) as well as non-homogeneous (different service time distributions and basestock levels) stages were evaluated. In addition, tandem systems that represented a mix of make-to-stock and make-to-order systems were also evaluated.

OBJECTIVE 2: The objective was to extend the basic framework developed as part of Objective 1 to model additional manufacturing features within the individual stages in the tandem configuration. Each sub-objective below addresses a specific manufacturing feature.

Sub-objective 2.1: The objective was to extend the analysis framework to model parallel machines at a stage. The approach involved the use of the approximation developed by Whitt (1993) for GI/G/m queues.

Sub-objective 2.2: The objective was to extend the basic decomposition approach to address the batching feature within the domain of tandem make-to-stock systems. In some manufacturing systems, orders are not released until batches of them have accumulated. It is common to wait for orders to be batched when long setup times are needed.

Sub-objective 2.3: The objective was to relax the assumption of unlimited supply of raw materials. In the tandem line configurations described in the previous objectives, it was assumed that raw materials were always available. This assumption was relaxed in this objective by incorporating the feature of limited raw materials inventory within the system. The supplier is an integral part of a supply chain system, and these models could be very useful in analyzing a supply chain network.

Sub-objective 2.4: The objective was to model multiple-part types which are another essential feature of many manufacturing systems. In systems that produced different part types it was assumed that each part type had its own inventory of finished goods as well as intermediate semi-finished parts. The basic aggregation approach used in the queueing network analyzer (Whitt, 1983) was used in combination with the decomposition framework to analyze these systems.

Sub-objective 2.5: The objective was to extend the analysis framework to incorporate service disruptions within the performance model. Broadly speaking, two classes of failures were modeled. One class includes disruptions that are quite frequent and do not take a long time to fix. These are usually operation dependent. Examples are machine stoppages due to tool breakage or part jams. These disruptions do not cause any structural changes in the system. These disruptions can be incorporated within the performance models by modifying service times to include the effect of such disruptions.

The other type of failure is the one that occurs infrequently such as the major breakdown of a critical piece of equipment. Also, the repair in such instances might take several days or even weeks. In the event of this type of breakdown, an alternative machine may be used so that the production system is not shut down. This causes structural changes and the performance of the system is affected. The performability framework is well suited to analyze such situations.

Objectives pertaining to non-tandem configurations.

OBJECTIVE 3: The objective was to develop models for tree-structured or feed-forward production-inventory networks. This configuration resembles the multi-echelon inventory system studied by Svoronos and Zipkin (1991).

OBJECTIVE 4: The objective was to develop models to address systems with feedback. Production-inventory systems can have parts fed back due to a part failing inspection and requiring rework; a few example systems were evaluated and results are presented in Chapter 7.

3.3 RESEARCH SCOPE AND LIMITATIONS

The models developed in this research are suitable for discrete part manufacturing systems. However, the scope of the research will be limited by the following assumptions.

- The queues at the individual machines are always assumed to be infinite. Thus blocking issues are not addressed in this research.
- The models developed do not consider the simultaneous possession of multiple resources. An example is the requirement of a machine, a tool and an operator before processing can begin.
- Supply of intermediate (purchased) parts are not considered in this research.
- Assembly operations and material handling issues are outside the scope of this research.

CHAPTER IV

MODELING TANDEM MAKE-TO-STOCK SYSTEMS: POISSON ARRIVALS AND EXPONENTIAL PROCESSING TIMES

4.1 INTRODUCTION

We begin the development of our approach by starting with the simplest configuration, a tandem make-to-stock system with Poisson arrivals and exponential processing times. Lee and Zipkin (1992) analyzed these types of systems, and their procedure was discussed in Chapter 2. A new decomposition approach is developed in this chapter, which forms the basis for a performance analysis framework. Early versions of this chapter were the subject of two conference presentations (Sivaramakrishnan and Kamath, 1996 and 1997). One of these was based on a refereed proceedings paper (Sivaramakrishnan and Kamath, 1997). Extensions to include additional manufacturing features and other system configurations are presented in subsequent chapters.

The remainder of this chapter is organized as follows. The next section describes the tandem system, its dynamics and the assumptions made. The mathematical procedure is described in an algorithmic form in Section 4.3, and the numerical investigation is reported in Section 4.4. The last section briefly mentions the extension of the procedure for general arrivals and general service times, and sets the stage for Chapter 5.

4.2 SYSTEM DESCRIPTION

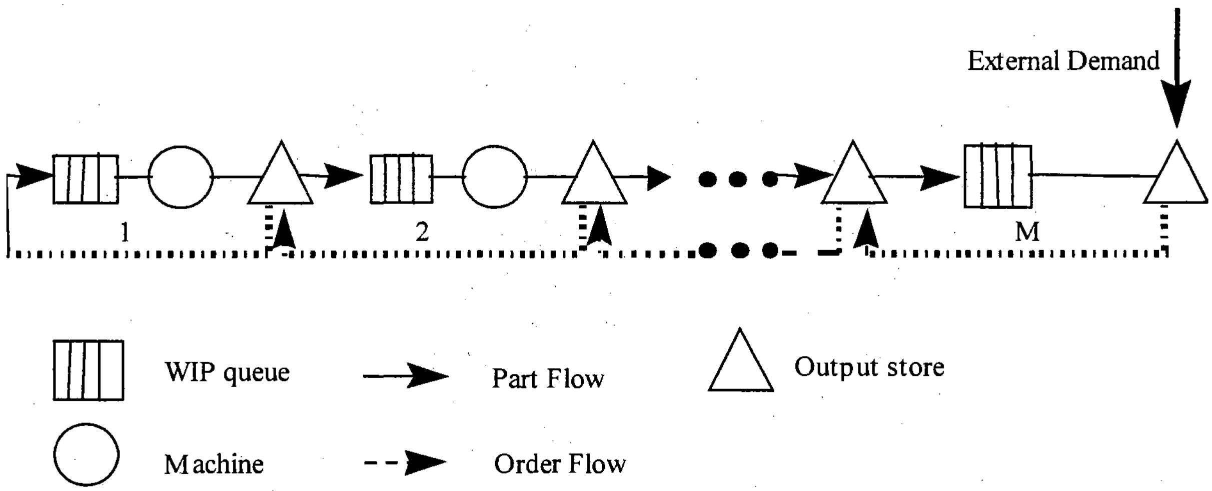


Figure 4. 1: An M-Stage Tandem Make-to-Stock Production System

Consider the M-stage tandem make-to-stock system shown in Figure 4.1. Processing begins at stage 1 and proceeds sequentially to stage M. It is assumed that at each stage there is a single server, representing a machine. Also, it is assumed that setup times are included in the processing times. The system is controlled by a stationary demand-pull or basestock policy. This policy is represented by the non-negative integers S_i, $i = 1, 2, ..., M$. The quantity S_i is called the basestock level at stage i, and it determines the maximum planned inventory of the output at stage i. The demand arrival process is Poisson and the processing times at each stage are exponentially distributed. Customer demand occurs at stage M and it is for one unit at a time. If finished goods are available then the demand is fulfilled immediately. An order for an item is triggered to replenish the finished goods inventory. This policy is termed as *one-for-one replenishment*. If finished goods are not available, the demand is backordered. The inventory is replenished until the basestock level S_M is reached. The order to replenish the finished goods stock looks into the output store of stage M-1. If parts are available at this store, the order picks up a part and proceeds for processing at stage M. If parts are unavailable, a backorder is logged at stage M-1. Likewise, each earlier stage fills its demand by releasing a unit to the next stage, if stock is available, and otherwise it logs a backorder. At stage 1, orders go immediately into the queue for processing, i.e., it is assumed that raw materials are always available. The queues where parts wait to be processed have unlimited capacities. Also, if there are outstanding backorders at a stage, the units that complete processing at this stage are released immediately to fulfill the backorders. A hybrid system consisting of make-to-stock stages and make-to-order stages can be modeled by constraining some of the basestock levels to be zero.

4.3 THE DECOMPOSITION APPROACH

In this section, a new decomposition procedure is developed and, all the required formulas for the analysis are presented. The following observation can be made about the system. It can be seen that if all the S_is are zero, the system becomes the classic tandem queue system, and the exact solution can be obtained using Jackson's (1957) results. This observation is not valid even if one of the S_is is greater than zero except in the case when all S_is, $i \neq M$ are zero and S_M is non-zero. Exact analysis of such systems becomes difficult in other cases.

The solution process begins with stage 1, which is an M/M/1 make-to-stock system because of our assumption that raw materials are always available. Using the formulas contained in Buzacott and Shanthikumar (1993) for a single stage make-to-stock system, all the steady state measures can be obtained for this stage as shown later in this section. There is no approximation needed for stage 1.

At stage 2, an order could be delayed because of the unavailability of parts in the output store of stage 1. In fact, this phenomenon is seen in all the remaining stages of the system. A modified single-stage system with a delay node (see Figure 4.2) is developed to handle this situation. In the system shown in Figure 4.2, an order goes to the delay node with a fixed probability p, before joining the processing queue. Using a procedure similar to that used by Buzacott and Shanthikumar (1993) for an M/M/1 make-to-stock system, the steady-state measures are derived for this system. This delay model is then used for the last M-1 stages of the M-stage make-to-stock system. *The delay node essentially captures the upstream delay experienced by an order when it does not find a part in the output store of the previous stage*. The analysis of the delay model is described in the next section.

4.3.1 A Single-Stage Model with a Delay Node

Consider the make-to-stock system shown in Figure 4.2. The demand for finished goods is Poisson with a rate λ. A one-for-one replenishment policy is followed; that is, every demand fulfilled from the output store triggers an order to replenish the finished goods inventory. The level of stock is a known quantity S. If there are no parts in stock, demand is backordered. With a fixed probability p, the orders for replenishment may be delayed by a random time with a mean τ_d, before joining the processing queue. The processing times are exponentially distributed with a rate of μ. The following notation is used.

I is the inventory level in the output store; B is the number of backorders in the system; and N is the number of orders (in delay + processing) in the system. The delay node could be viewed as an infinite server node with τ_d as the average service time. The average number of busy servers gives the average number at the delay node, which is $\rho_d = (\lambda.p).\tau_d$. The utilization at the processing node is given by $\rho = \dfrac{\lambda}{\mu}$.

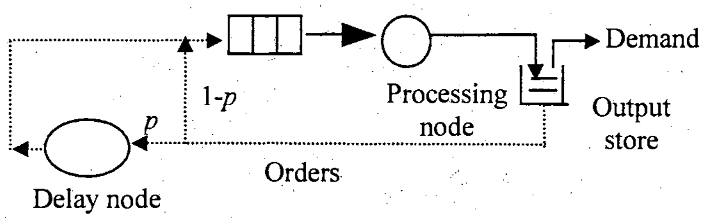


Figure 4. 2: A Single-Stage Make-to-Stock System with a Delay Node

Using standard product-form queueing network theory (Buzacott and Shanthikumar, 1993), we can find the steady state probability that there are n orders in the system.

$$
\begin{aligned}
P[N = n] &= \sum_{i=0}^{n} \text{P}[i \text{ orders at the processing node}].\text{P}[n\text{-}i \text{ orders at the delay node}] \\
&= \sum_{i=0}^{n} (1-\rho).\rho^{i}.\frac{e^{-\rho_d}.(\rho_d)^{n-i}}{(n-i)!} \quad . \qquad (4.1) \\
&= e^{-\rho_d}.(1-\rho)\sum_{i=0}^{n} \frac{\rho_d^{n-i}.\rho^{i}}{(n-i)!}, n = 0,1,2,\ldots
\end{aligned}
$$

Using the above expression and a procedure similar to that used by Buzacott and Shanthikumar (1993) for a single-stage make-to-stock system, we have for inventory level, I,

$$
\begin{aligned}
P[I = k] &= P[N = S - k]; k = 1,2,\cdots,S \\
&= P[N \geq S]; k = 0
\end{aligned} \qquad (4.2)
$$

Hence,

$$P[I=k]=e^{-\rho_d}.(1-\rho)\sum_{i=0}^{S-k}\frac{\rho_d^{S-k-i}.\rho^i}{(S-k-i)!};k=1,2,\cdots,S \quad (4.3)$$

$$P[I=0]=\sum_{k=S}^{\infty}P[N=k] \quad (4.4)$$

The average inventory in stock is simply

$$E[I]=\sum_{k=1}^{S}k.P[I=k] \quad (4.5)$$

The steady-state probability that a demand will not find a part in the output store is given by the steady-state probability that the output store is empty, which is $1-\sum_{k=1}^{S}P[I=k]$. This is a consequence of the PASTA (Poisson Arrivals See Time Averages) property (Wolff, 1982). The expected number of backorders in the system is given by the following relationship.

$$E[B]=E[N]+E[I]-S \quad (4.6)$$

4.3.2 Analysis of the M-Stage Line

Beginning at stage 2, each of the remaining M-1 stages is modeled as a single-stage make-to-stock system with a delay node. In other words, the M-stage tandem make-to-stock system is decomposed into one single-stage system plus M-1 single-stage systems each with a delay node (see Figure 4.3). The delay node at each of the stages captures the upstream delay which occurs when an order waits for a part from the previous stage. This procedure is done sequentially beginning at stage 1. At each stage other than the first stage, we need to know the probability that an order proceeds to the delay node, that is, it is backordered. This probability is same as the probability that an order will not find a part in the output store of stage *i*-1. Also, the average number at the delay node is the average number of backorders at stage *i*-1. Using these observations, we present the procedure in a step-wise manner with the required mathematical expressions.

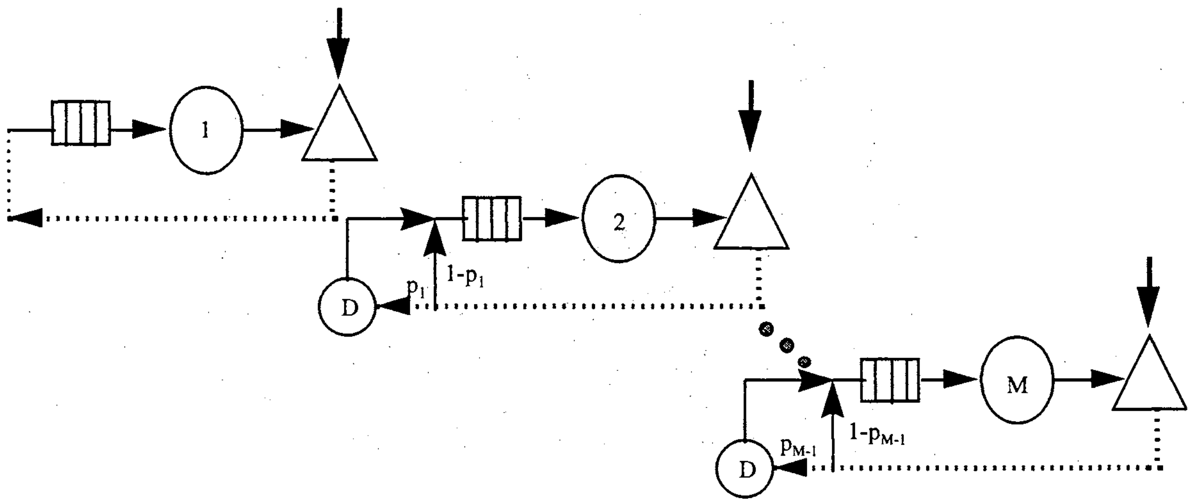


Figure 4. 3: Decomposition of an M-Stage Make-to-Stock System into M Single-Stage Make-to-Stock Systems

The parameters of the model are:

M = number of stages;

λ = the demand rate;

μ_i = the service rate at stage i, $i = 1, 2, ..., M$; and

S_i = the basestock level at stage i, $i = 1, 2, ..., M$.

The performance measures used in evaluating the system are as follows.

ρ_i = λ / μ_i is the utilization at stage i;

$E[B_i]$ = the average number of backorders at stage i;

$E[I_i]$ = the average inventory in the output store at stage i;

E[N_i] = the average number at stage i (including the one in processing); and

p_i = the probability that there is no part in the output store of stage i when a request by an order from the stage i+1 is made, i = 1, 2, ..., M-1.

Stage 1:

The expected number in system and the expected number of backorders in this stage is given by,

$$E[N_1] = \frac{\rho_1}{1-\rho_1} \tag{4.7}$$

$$E[B_1] = \frac{\rho_1^{S_1+1}}{1-\rho_1} \tag{4.8}$$

Also, $p_1 = \rho^{S_1}$.

The above measures are obtained from the analysis of an M/M/1 make-to-stock system (Buzacott and Shanthikumar, 1993).

Stage *i* (*i* > 1): The average number at node i is the sum of the orders waiting for parts from the previous stage and the orders with parts waiting to complete processing (including the one in process). Hence,

$$E[N_i] = E[B_{i-1}] + \frac{\rho_i}{1-\rho_i} \tag{4.9}$$

For simplicity, we do not add more subscripts to λ_d, τ_d and ρ_d.

The arrival rate to the delay node is given by $\lambda_d = \lambda . p_{i-1}$.

Using Little's (1961) result, the average time a part spends at the delay node; $\tau_d = \dfrac{E[B_{i-1}]}{\lambda_d}$.

Now, $\rho_d = \lambda_d \tau_d$. Note that ρ_d is simply the average number of backorders at stage i-1. Using a single-stage model with delay node, the steady state probability that there are k parts in the output store is given by,

$$P[I_i = k] = e^{-\rho_d}(1-\rho_i)\sum_{j=0}^{S_i-k}\frac{\rho_d^{S_i-k-j}}{(S_i-k-j)!}\rho_i^j \text{ , } k = 1, 2, ..., S_i. \tag{4.10}$$

The expected inventory at stage i is

$$E[I_i] = \sum_{n=1}^{S_i} k.P[I_i = k]. \tag{4.11}$$

Using (4.9) and (4.11), the expected backorders at stage i is given by,

$$E[B_i] = E[N_i] + E[I_i] - S_i \tag{4.12}$$

The probability that a demand from stage i+1 will not find a part in the output store of stage i is

$$p_i = 1 - \sum_{n=1}^{S_i} P[I_i = n] \tag{4.13}$$

Beginning at Stage 2, this procedure is repeated sequentially till the last stage.

4.4 NUMERICAL RESULTS

In this section we present results obtained using our decomposition method for some example systems analyzed by Lee and Zipkin (1992). As mentioned earlier, Lee and Zipkin

(1992) was one of the first published journal article that examined tandem make-to-stock systems. Lee and Zipkin (1992) set λ =1 for all the systems they analyzed. They analyzed a variety of two-stage and three-stage systems. Our results match exactly with results obtained by Lee and Zipkin's method for all the two-stage systems examined by them.

In the three-stage systems, Lee and Zipkin (1992) (L&Z) restrict their attention to systems where the service rates are equal at all the stages. They examined systems with μ = 1.25, 1.5, and 2. Also, they used two values for S_1 and S_2, viz., 3 and 7, and obtained simulation estimates for all possible combinations. In total, they obtained estimates for 12 different cases. In all the cases, S_3 was set to zero. Table 4.1 presents the average backorders at stage 3 and Table 4.2 presents the average intermediate inventory which is the sum of inventory at the output stores of stages 1 and 2 and work-in-process in Stages 2 and 3. The parts waiting in queue and in process at Stage 1 are considered to be new material and hence, not included in the average intermediate inventory calculation.

Table 4. 1: Estimates of E[B3] (3-Stage System)

μ	S_1	S_2	Simulation L&Z	L&Z method	Our Method	% difference (L&Z)	% difference (Ours)
1.25	3	3	7.284	7.726	7.385	6.06	1.38
1.25	3	7	5.523	5.870	5.400	6.28	-2.22
1.25	7	3	6.463	6.735	6.525	4.22	0.95
1.25	7	7	5.046	5.261	5.035	4.27	-0.22
1.50	3	3	2.690	2.944	2.802	9.45	4.16
1.50	3	7	2.162	2.233	2.161	3.26	-0.04
1.50	7	3	2.537	2.662	2.633	4.93	3.78
1.50	7	7	2.110	2.140	2.128	1.40	0.85
2.00	3	3	1.120	1.164	1.142	3.94	1.96
2.00	3	7	1.001	1.012	1.009	1.08	0.80
2.00	7	3	1.089	1.127	1.126	3.49	3.40
2.00	7	7	1.014	1.008	1.008	0.62	0.62

Table 4. 2: Average Intermediate Inventory (3-Stage System)

μ	S_1	S_2	Simulation L&Z	L&Z method	Our Method	% difference (L&Z)	% difference (Ours)
1.25	3	3	9.288	9.726	9.385	4.72	1.04
1.25	3	7	11.517	11.870	11.400	3.07	-1.01
1.25	7	3	12.461	12.735	12.525	2.20	0.51
1.25	7	7	15.060	15.261	15.035	1.33	-0.17
1.50	3	3	6.722	6.944	6.799	3.30	1.14
1.50	3	7	10.151	10.233	10.158	0.81	0.06
1.50	7	3	10.550	10.662	10.631	1.06	0.76
1.50	7	7	14.101	14.140	14.125	0.28	0.17
2.00	3	3	6.722	6.142	6.164	1.96	0.68
2.00	3	7	10.004	10.009	10.012	0.80	0.08
2.00	7	3	10.092	10.550	10.127	3.40	0.35
2.00	7	7	14.006	14.101	14.008	0.62	0.02

The results are not exact since the analytical method assumes that the arrival distribution at each stage is Poisson. The arrival of demand at an output store is Poisson, but the arrival of orders into the processing queue at a node (except node 1) is not a Poisson process. This is because of the delay experienced by some of the orders before they proceed to the processing queue. The delay nodes approximately capture the inter-dependence between the stages.

The results indicate that our method performs better than Lee and Zipkin's approximation in all the cases for the two performance measures examined. In all the cases examined, the relative percentage difference was less than 5% and it is also recognized that the model by Lee and Zipkin (1992) performed reasonably well in all the cases examined.

4.5 SUMMARY

In this chapter, a new decomposition framework for the analysis of tandem make-to-stock systems with Poisson arrivals and exponential processing times was developed. As shown in the subsequent chapters, this framework was also used for the analysis of more complex configurations. In the next chapter, we describe how we can handle general demand processes and general processing time distributions within the decomposition framework. Conceptually, the approximation approach remains the same. We use two-moment approximations that require only the mean and squared coefficient of variation of the inter-arrival and service time distributions (Whitt, 1983; Segal and Whitt, 1989) in place of the exponential queueing models. The analysis procedure for the M-stage line proceeds in a manner similar to that described in Section 4.2. Further details are contained in the next chapter. Following the next chapter, extensions to include multiple servers, multiple part types, batch service and breakdowns of machines are presented.

CHAPTER V

MODELING TANDEM MAKE-TO-STOCK SYSTEMS: GENERAL ARRIVALS AND GENERAL PROCESSING TIMES

5.1 OVERVIEW

The models discussed in the previous chapter assumed that the demand arrived according to Poisson process and that the processing times were exponential distributed. In this chapter, these assumptions are relaxed, and methods are developed to analyze tandem systems with a general demand arrival process and general processing times. Conceptually, the decomposition approach remains the same. The queueing analysis is now based on two-moment approximations that require only the mean and squared coefficient of variation (SCV) of the inter-arrival and service time distributions (Whitt, 1983; Segal and Whitt, 1989).

The first stage now becomes a GI/G/1 make-to-stock system, and the steady-state formulas given by Buzacott and Shanthikumar (1993) are used. For a single-stage system with a delay node, we first solve for the steady-state probability of number in system using the parametric decomposition method outlined in Whitt (1983). The arrival process to a processing stage is now the result of the merging of two arrival streams - parts arriving from the previous stage to fulfill backorders at the current stage and orders that find parts (at the output store of the previous stage) and proceed immediately to the processing queue. The departure process leaving a stage is split into two streams - one that proceeds to the next stage to satisfy backorders and another that goes to the output store to satisfy replenishment orders. The next few sections contain the details of the approximation scheme. This is followed by a presentation of extensive numerical results for several example systems.

5.2 MODEL DESCRIPTION

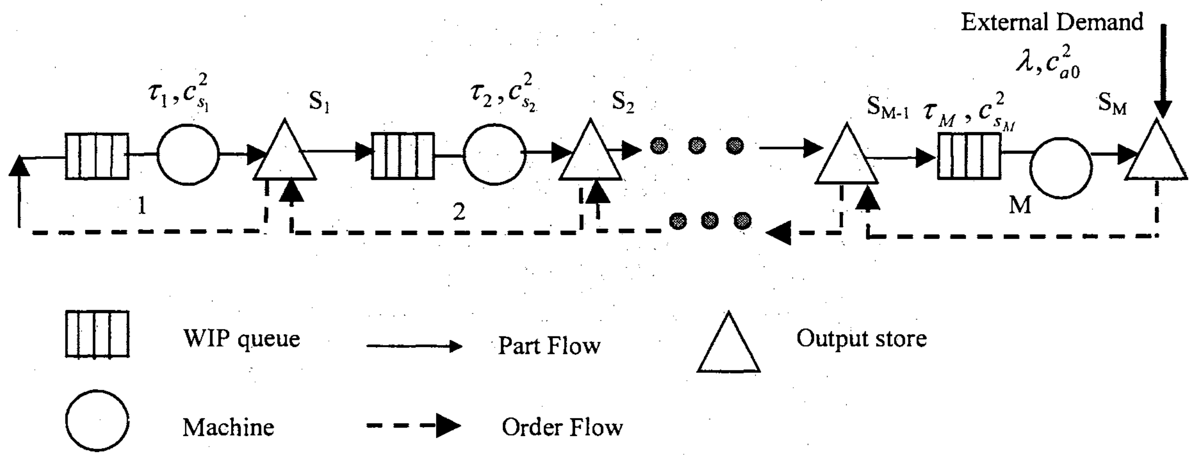


Figure 5. 1: A General Tandem Make-to-Stock System

Consider the M-stage tandem make-to-stock system shown in Figure 5.1. This system is similar to that described in Section 4.2. The major differences are that the demand process is a renewal process, and the processing times at each stage are generally distributed. As before the replenishment policy is one-for-one, with the quantity S_i representing the basestock level at stage i, $i = 1, 2, \ldots, M$.

5.3 GENERALIZATION OF THE APPROXIMATION SCHEME

Conceptually, the approximation scheme is similar to that developed for the exponential case. Hence, we focus only on the generalization aspects here. Stage 1 is now a GI/G/1 make-to-stock system. Using the approximations contained in Buzacott and Shanthikumar (1993), all of the steady state measures can be obtained for stage 1 as shown in Section 5.3.3.

At stage 2, an order may be delayed before proceeding to get processed because of the unavailability of parts in the output store of stage 1, which is seen in all the remaining stages of the system. Again, we use a modified single-stage system with a delay node (see Figure 5.2) to handle this situation. The delay node (an infinite server system) essentially captures the upstream delay when a order does not find a part in the output store of the previous stage. The analysis of the delay model is described next.

5.3.1 A General Single-Stage Model with a Delay Node

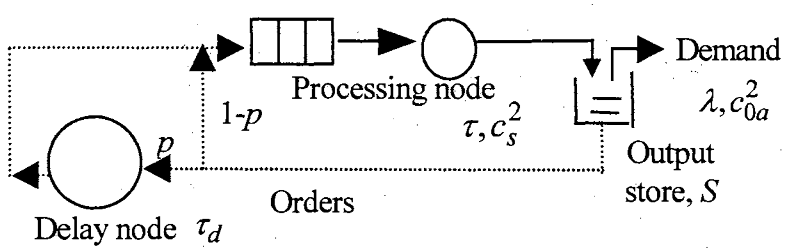


Figure 5. 2: A General Single-Stage Make-to-Stock System with a Delay Node

Consider the make-to-stock system shown in Figure 5.2. The demand for finished goods is a renewal process with a rate λ, and the squared coefficient of variation (SCV) of inter-arrival times, c_{a0}^2. We have a stock of finished goods in the output store, which has a capacity of S units. A one-for-one replenishment policy is followed. If there are no parts in stock, demand is backordered. With a probability p, the orders for replenishment may be delayed by a random time with a mean τ_d before getting processed. The processing times follow a general distribution with a mean of τ and SCV, c_s^2. The following notation is used.

I = the inventory level in the output store;

B = the number of backorders in the system;

N = the number of orders (in delay + processing) in the system;

N_d = the number of orders at the delay node; and

N_p = the number of orders at the processing node.

We model the delay node by an infinite server system. Now, the average number at the delay node is given by the average number of busy servers, which is $\rho_d = \lambda.p.\tau_d$ (=E[N_d]). We approximate the distribution of the number in process at the delay node using the M/G/∞ formula which is given by $P[N_d = k] = \dfrac{e^{-\rho_d}(\rho_d)^k}{k!}$. The utilization at the processing node is given by $\rho_p = \lambda.\tau$.

We then find the approximate steady state probability that there are n orders in the system, by assuming that the delay and processing nodes behave like independent, isolated nodes as in the product-form case (Whitt, 1983).

$$P[N = n] \cong \sum_{j=0}^{n} P[N_p = j].P[N_d = n - j] \tag{5.1}$$

$$P[N = n] \cong \begin{cases} (1-\rho_p).\dfrac{e^{-\rho_d}(\rho_d)^n}{n!} + \displaystyle\sum_{j=1}^{n} \rho_p(1-\sigma)\sigma^{j-1}.\dfrac{e^{-\rho_d}(\rho_d)^{n-j}}{(n-j)!} & n = 1, 2, \ldots \\ e^{-\rho_d}.(1-\rho_p) & n = 0 \end{cases} \tag{5.2}$$

where $\sigma = \frac{E[N_p] - \rho_p}{E[N_p]}$

E[N_p] denotes the expected number at the processing node which is calculated using the GI/G/1 formulas from Whitt (1983) as shown below. We first use the Kraemer and Langenbach-Belz (1976) approximation to calculate the expected waiting time in queue at the processing node

$$E[W_q] \cong g\left(\frac{c_a^2 + c_s^2}{2}\right)\left(\frac{\tau\rho_p}{(1-\rho_p)}\right) \tag{5.3}$$

where $g \equiv g(\rho_p,\, c_a^2,\, c_s^2)$ is defined as

$$g(\rho_p,\, c_a^2,\, c_s^2) = \begin{cases} \exp\left(\frac{-2(1-\rho_p)(1-c_a^2)^2}{3\rho_p(c_a^2 + c_s^2)}\right), & c_a^2 < 1 \\ \exp\left(\frac{-(1-\rho_p)(c_a^2 - 1)}{(c_a^2 + 4c_s^2)}\right), & c_a^2 > 1 \end{cases} \tag{5.4}$$

where c_a^2 is the SCV of the inter-arrival times at the processing node.

We know all the parameters except c_a^2. A procedure to directly calculate c_a^2 is presented in the next section. When the delay model is used repeatedly in the sequential solution algorithm, we have enough information to calculate c_a^2. Hence, we will proceed here by assuming that c_a^2 is available. Using Little's law, the expected number at processing node is given by,

$$E[N_p] = \lambda . E[W_q] + \rho_p \tag{5.5}$$

The steady-state probability that there are k parts in the output store is given by $P[N = S - k]$ for $0 < k < S$. Hence,

$$P[I = k] \cong \begin{cases} e^{-\rho_d}(1-\rho_p)\dfrac{\rho_d^{S-k}}{(S-k)!} + \displaystyle\sum_{j=1}^{S-k} \frac{e^{-\rho_d}\rho_d^{S-k-j}}{(S-k-j)!}\rho_p.\sigma^{j-1}(1-\sigma) & k = 1, 2, ..., S-1 \\ e^{-\rho_d}.(1-\rho_p) & k = S \end{cases} \tag{5.6}$$

The average inventory in stock is simply

$$E[I] = \sum_{k=1}^{S} n.P[I = k] \tag{5.7}$$

The average number in the system, $E[N] = E[N_d] + E[N_p]$.

The expected number of backorders in the system is given by the following relationship (Lee and Zipkin, 1992).

$$E[B] = E[N] + E[I] - S \tag{5.8}$$

The steady-state probability, p, that a demand will not find a part in the output store is approximated by the steady-state probability that the output store is empty, which is

$$1 - \sum_{k=1}^{S} P[I = k]. \tag{5.9}$$

5.3.2 Determining the SCV of the Combined Arrival Process at the Processing Node

In order to determine the SCV of the arrival process at the processing node, we shift our focus to examine the splitting of the departure process and merging process of arrival processes between any two successive stages in the multi-stage, make-to-stock system. Figure 5.3 shows these processes between stages i and i+1.

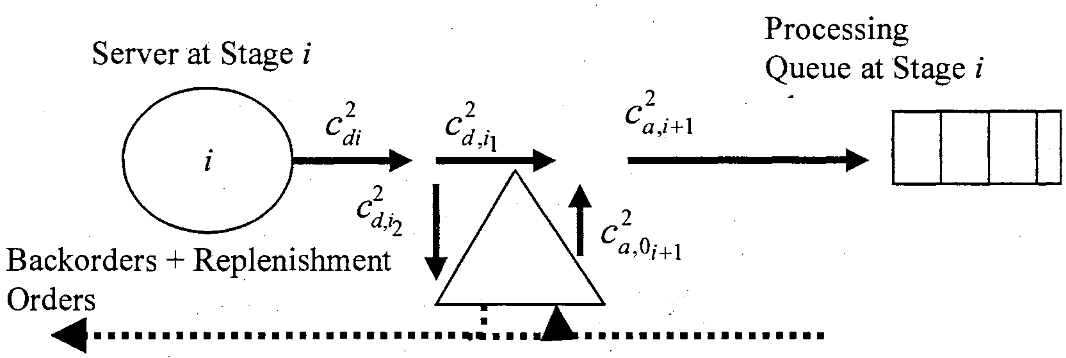


Figure 5. 3: Splitting and Merging Processes between Stages

At stage i, the departure process splits into two, namely, the process which satisfies the backorders at stage $i+1$ and the process which satisfies the replenishment orders at the output store of stage i. Focusing on the arrival process at stage $i+1$, it is the merging of the two processes namely, those orders which find parts at the output store of stage i and directly proceed to stage $i+1$, and those which are backordered indicated by the parts which proceed directly from the stage i to stage $i+1$. The arrival rates at both the nodes are the same as the external arrival rate because the external arrival process is the only process that triggers orders and we have assumed a one-for-one replenishment policy. We use the following procedure to determine the SCV of the arrival process at each stage. It should also be noted that all the performance measures pertaining to stage i are known at this juncture.

Let c_{ai}^2 be the SCV of the inter-arrival distribution at stage i, c_{si}^2 the SCV of the service times at stage i, and $c_{a0_{i+1}}^2$ the SCV of the interval times of the external arrivals that go directly into stage $i+1$. p_i is the steady-state probability that a demand from stage $i+1$ will be backordered at stage i.

Using the splitting approximation from Whitt (1983), we have

$$c_{a0_{i+1}}^2 \cong (1-p_i)c_{a0}^2 + p_i \tag{5.9}$$

The departure SCV at node i is given by (Whitt, 1983)

$$c_{di}^2 \cong \rho_i^2 c_{si}^2 + (1-\rho_i^2)c_{ai}^2 \tag{5.10}$$

Again, using the splitting approximation from Whitt (1983), SCV of the process that goes into stage $i+1$ is given by

$$c_{d,i1}^2 = p_i c_{di}^2 + 1 - p_i \tag{5.11}$$

where the p_i is the probability that the completed part satisfies a backorder. This is also the probability that an arrival was backordered at the output store of stage i.

Using the method of superposition from Whitt (1983), the SCV of the arrival process at stage i+1 is

$$c_{a,i+1}^2 \cong p_i c_{di1}^2 + (1-p_i).c_{a0_{i+1}}^2 \tag{5.12}$$

The SCV thus determined is used as the SCV of the arrival process at the processing node in the delay model described in the previous section.

5.3.3 Analysis of the General M-Stage Tandem Line

As in Section 4.3.2, we decompose the M-stage tandem make-to-stock system into one single-stage system plus M-1 single-stage systems each with a delay node (see Figure 5.4).

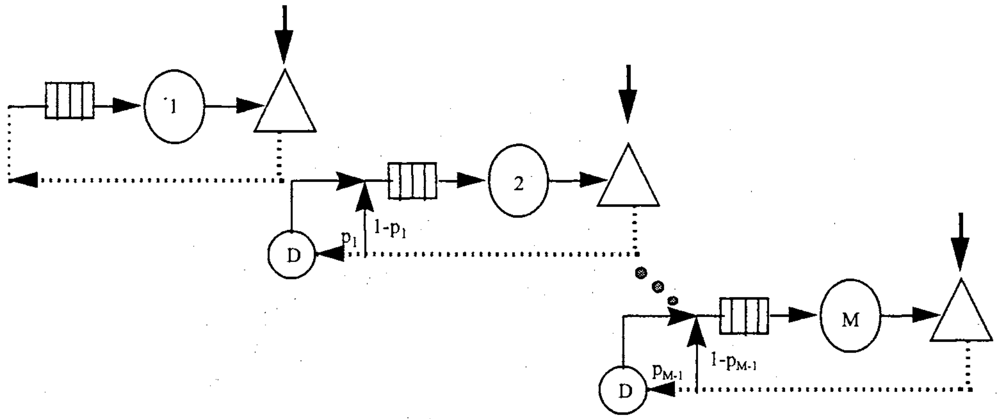


Figure 5. 4: Decomposition of an M-Stage Make-to-Stock System into M Single Make-to-Stock Systems

We now present the analysis algorithm for the general case.

The parameters of the model are:

M = number of stages;

λ = the demand arrival rate;

τ_i = the mean processing time at stage i;

c_{a0}^2 = the SCV of the inter-arrival time of the demand process;

c_{si}^2 = the SCV of the processing time at stage i; and

S_i = the basestock level at stage i.

Some intermediate quantities are as follows.

ρ_i = $\lambda\tau_i$ is the utilization at stage i; and

p_i = the probability that there is no part in the output store of stage i when a request is made by an order from the stage i+1;

The node measures used in evaluating the system performance are as follows.

$E[B_i]$ = the average number of backorders at stage i;

$E[I_i]$ = the average inventory in the output store at stage i; and

$E[N_i]$ = the average number at stage i (including the one in processing).

In the above quantities, $i = 1, 2, \ldots, M$.

Stage 1:

The expected number of orders and the expected number of backorders in this stage are given by,

$$E[N_1] = \lambda . E[W_{q1}] + \rho_1; \; E[B_1] = \frac{\rho_1 \sigma_1^{S_1}}{(1-\sigma_1)}; \text{ where } \sigma_1 = \frac{E[N_1] - \rho}{E[N_1]} \tag{5.13}$$

$E[W_{q1}]$ is the expected waiting time in queue at stage 1 and is calculated using the GI/G/1 approximation given in Section 5.3.1. The expression for $E[B_1]$ is obtained from the analysis of a GI/G/1 make-to-stock system contained in Buzacott and Shanthikumar (1993).

The probability that a demand from stage 2 will not find a part in the output store of stage 1 is approximated by the steady state probability that the output store of stage 1 is empty. That is

$$p_1 \cong 1 - \sum_{k=1}^{S_1} P[I_1 = k] \tag{5.14}$$

Stage *i* (*i* > 1):

The SCV of the arrival process c_{ai}^2 is calculated using the procedure given in Section 5.3.2. The average number at node i is the sum of the orders waiting for parts from the previous stage, stage i-1, and the orders with parts waiting to complete processing (including the one in process). Hence,

$$\mathrm{E}[\mathrm{N}_i] = \mathrm{E}[\mathrm{B}_{i-1}] + \lambda.\mathrm{E}[\mathrm{W}_{qi}] + \rho_i \tag{5.15}$$

For simplicity, we do not add more subscripts to λ_d, τ_d and ρ_d. The arrival rate to the delay node is given by

$$\lambda_d = p_{i-1}.\lambda \tag{5.16}$$

The average time a part spends at the delay node is

$$\tau_d = \frac{\mathrm{E}[\mathrm{B}_{i-1}]}{\lambda_d} \tag{5.17}$$

Now, $\rho_d = \lambda_d \tau_d$. Note that ρ_d is simply the average number of backorders at stage i-1. Using the single-stage model with delay node, the steady state probability that there are k parts in the output store is given by,

$$P[I_i = k] \cong \begin{cases} e^{-\rho_d} \dfrac{\rho_d^{S_i-k}}{(S_i-k)!}(1-\rho_i) + \displaystyle\sum_{j=1}^{S_i-k} e^{-\rho_d} \dfrac{\rho_d^{S_i-k-j}}{(S_i-k-j)!}\rho_i.(1-\sigma_i).\sigma_i^{j-1} & k = 1,2,...,S_i-1 \\ e^{-\rho_d}.(1-\rho_i) & k = S_i \end{cases} \tag{5.18}$$

The expected inventory at stage i is

$$E[I_i] = \sum_{k=1}^{S_i} k.P[I_i = k] \tag{5.19}$$

The expected backorders at stage i is given by,

$$\mathrm{E}[\mathrm{B}_i] = \mathrm{E}[\mathrm{N}_i] + \mathrm{E}[\mathrm{I}_i] - \mathrm{S}_i \tag{5.20}$$

Again, the probability that a demand from stage i+1 will not find a part in the output store of stage i is approximated by

$$p_i \cong 1 - \sum_{k=1}^{S_i} P[I_i = k] \tag{5.21}$$

Beginning at Stage 2, this procedure is repeated sequentially till the last stage.

5.4 NUMERICAL RESULTS

In this section, an investigation of the accuracy of the approximation is presented, by comparing its predictions to estimates from computer simulation experiments. First, we examine a variety of homogeneous three-stage systems wherein the parameters at each stage are the same. In other words, the utilization or mean service time, the basestock level and the SCV of service time distribution are the same across all the stages in a particular configuration.

5.4.1 Homogeneous Systems

We set λ =1 for all the systems. Three different values of the ρ, viz., 0.6, 0.7 and 0.8, are used in combination with three different basestock levels, S = 3, 6 and 9. The inter-arrival and service time distribution SCVs used are 0.25, 1, and 2.25; these correspond to the Erlang, exponential and hyper-exponential distributions, respectively. The performance measures of interest are the average backorders at the last stage, E[B_3], the average inventory level at the last stage, E[I_3], and the average intermediate inventory, $(\sum_{i=1}^{2} \mathrm{E}[\mathrm{I}_i] + \mathrm{E}[\mathrm{N}_{i+1}])$, where E[$N_i$] is the expected number in the processing node at stage i.

Simulation estimates were obtained using a program developed in SLAM II (Pritsker, 1995). The method of replication deletion was used to obtain statistically accurate simulation estimates. Appropriate values for warm-up, run length and the number of replications were determined for the system with the highest variability and were used across all the configurations. The warm-up analysis was performed using Welch's (1983) procedure. A warm-up period of 5,000 time units, a run length of 50,000 time units after warm-up and 10 replications were used for a single configuration. Additional details about the warm-up and run length determination are presented in Appendix A1. We present the results for the various configurations tested using tables and graphs.

Tables 5.1 through 5.9 show the estimates of $E[B_3]$, $E[I_3]$ and average intermediate inventory for various combinations of inter-arrival time and processing time distributions. Figures 5.5 through 5.11 graphically show the comparison between the analytical and simulation estimates for some of the configurations. Table 5.10 presents results for a ten-stage system. The results indicate that the performance of the analytical models in case of ten stage systems is similar to that of the three stage systems. In general, the difference in the estimates of average number of backorders at the last stage is wider in configurations when hyper-exponential distribution was chosen for either the arrival or the service process. A measure that is often used to evaluate the accuracy of the analytical model is the

$$\text{Relative percent difference (RPD)} = \frac{\text{Analytical Result - Simulation Estimate}}{\text{Simulation Estimate}} * 100$$

The RPD for the average intermediate inventory measure is within the acceptable range (< 15%) in most cases. Overall, the approximation performs reasonably well (RPD < 12%) in most cases, and performs very well (RPD < 8%) in cases in which the squared coefficients of variation of both the inter-arrival and service times are less than or equal to one. Appendix A2 presents a variation of the approximation approach that produces better results for the high SCV (hyper-exponential) cases.

Table 5. 1: Erlang Inter-Arrival Times and Erlang Processing Times

ρ	Basestock Level	Average Backorders at Stage 3		Average Inventory at Stage 3		Average Intermediate Inventory	
		Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
0.60	3	0.003	0.007	2.257	2.242	6.003	6.029
0.60	6	0.000	0.000	5.259	5.263	11.999	12.000
0.60	9	0.000	0.000	8.258	8.264	17.999	18.000
0.70	3	0.020	0.067	2.004	1.887	6.011	6.184
0.70	6	0.000	0.001	4.995	5.001	11.980	12.003
0.70	9	0.000	0.000	7.997	8.004	18.001	18.000
0.80	3	0.153	0.592	1.580	1.164	6.103	6.964
0.80	6	0.009	0.017	4.517	4.486	12.021	12.068
0.80	9	0.001	0.001	7.530	7.533	17.987	18.005

Table 5. 2: Erlang Inter-Arrival Times and Exponential Processing Times

ρ	Basestock Level	Average Backorders at Stage 3		Average Inventory at Stage 3		Average Intermediate Inventory	
		Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
0.60	3	0.108	0.172	1.960	1.829	6.097	6.283
0.60	6	0.006	0.008	4.965	4.929	12.000	12.018
0.60	9	0.000	0.001	7.971	7.939	17.995	18.001
0.70	3	0.521	0.746	1.449	1.180	6.505	6.968
0.70	6	0.048	0.068	4.433	4.330	12.075	12.140
0.70	9	0.006	0.010	7.145	7.390	17.993	18.022
0.80	3	2.433	3.149	0.736	0.340	8.116	9.153
0.80	6	0.484	0.642	3.357	2.980	12.521	13.006
0.80	9	0.128	0.146	6.336	6.196	18.197	18.295

Table 5. 3: Erlang Inter-Arrival Times and Hyper-Exponential Processing Times

ρ	Basestock Level	Average Backorders at Stage 3		Average Inventory at Stage 3		Average Intermediate Inventory	
		Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
0.60	3	0.840	0.831	1.456	1.257	6.846	6.956
0.60	6	0.135	0.130	4.413	4.316	12.176	12.196
0.60	9	0.024	0.016	7.445	7.366	18.010	18.044
0.70	3	3.066	2.800	0.795	0.469	8.690	8.716
0.70	6	0.771	0.678	3.353	3.166	12.859	12.898
0.70	9	0.006	0.202	7.445	6.281	17.994	18.307
0.80	3	9.387	9.751	0.256	0.020	13.552	14.078
0.80	6	4.496	3.798	1.722	1.137	16.134	16.009
0.80	9	1.896	1.556	4.297	3.909	19.945	19.994

Table 5. 4: Poisson Arrivals and Erlang Processing Times

ρ	Basestock Level	Average Backorders at Stage 3		Average Inventory at Stage 3		Average Intermediate Inventory	
		Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
0.60	3	0.135	0.153	1.938	1.843	6.036	6.147
0.60	6	0.012	0.015	4.854	4.838	12.005	12.015
0.60	9	0.001	0.002	7.830	7.838	17.999	18.002
0.70	3	0.416	0.466	1.572	1.329	6.136	6.416
0.70	6	0.048	0.079	4.433	4.280	12.047	12.078
0.70	9	0.012	0.016	7.299	7.279	18.005	18.016
0.80	3	1.480	1.515	1.057	0.586	6.589	7.130
0.80	6	0.410	0.429	3.507	3.229	12.069	12.401
0.80	9	0.128	0.143	6.298	6.203	18.032	18.410

Table 5. 5: Poisson Arrivals and Exponential Processing Times

ρ	Basestock Level	Average Backorders at Stage 3		Average Inventory at Stage 3		Average Intermediate Inventory	
		Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
0.60	3	0.518	0.424	1.656	1.521	6.370	6.402
0.60	6	0.084	0.047	4.514	4.500	12.066	12.073
0.60	9	0.015	0.015	7.513	7.500	18.008	18.015
0.70	3	1.580	1.293	1.143	0.833	7.133	7.126
0.70	6	0.573	0.313	3.611	3.671	12.342	12.309
0.70	9	0.114	0.098	6.685	6.667	18.098	18.098
0.80	3	5.049	4.524	0.581	0.139	9.408	9.385
0.80	6	1.816	1.474	2.616	2.111	13.233	13.363
0.80	9	0.786	0.626	5.222	5.012	18.643	18.614

Table 5. 6: Poisson Arrivals and Hyper-Exponential Processing Times

ρ	Basestock Level	Average Backorders at Stage 3		Average Inventory at Stage 3		Average Intermediate Inventory	
		Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
0.60	3	1.673	1.142	1.260	1.039	7.368	7.041
0.60	6	0.426	0.297	3.965	3.939	12.392	12.295
0.60	9	0.126	0.097	6.940	6.937	18.138	18.097
0.70	3	4.696	3.542	0.703	0.301	9.637	8.887
0.70	6	1.767	1.108	2.855	2.681	13.466	13.072
0.70	9	0.705	0.463	5.689	5.646	18.629	18.463
0.80	3	12.708	11.608	0.260	0.006	15.502	14.602
0.80	6	6.933	5.061	1.438	0.703	17.570	16.359
0.80	9	4.062	2.424	3.423	3.096	21.600	20.328

Table 5. 7: Hyper-Exponential Inter-Arrival Times and Erlang Processing Times

ρ	Basestock Level	Average Backorders at Stage 3		Average Inventory at Stage 3		Average Intermediate Inventory	
		Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
0.60	3	0.744	0.334	1.638	1.590	6.151	6.180
0.60	6	0.205	0.080	4.243	4.475	12.016	12.041
0.60	9	0.067	0.020	7.120	7.445	17.998	18.010
0.70	3	1.695	0.860	1.252	0.962	6.425	6.379
0.70	6	0.641	0.309	3.544	3.672	12.045	12.118
0.70	9	0.269	0.123	6.277	6.556	18.028	18.048
0.80	3	4.144	2.579	0.817	0.220	7.210	6.856
0.80	6	2.055	1.060	2.639	2.342	12.359	12.215
0.80	9	1.147	0.629	4.907	4.987	18.099	18.139

Table 5. 8: Hyper-Exponential Inter-Arrival Times and Exponential Processing Times

ρ	Basestock Level	Average Backorders at Stage 3		Average Inventory at Stage 3		Average Intermediate Inventory	
		Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
0.60	3	1.507	0.682	1.395	1.272	6.805	6.460
0.60	6	0.477	0.203	3.935	4.126	12.212	12.127
0.60	9	0.162	0.069	6.798	7.076	18.024	18.042
0.70	3	3.584	1.930	0.940	0.530	8.003	7.201
0.70	6	1.451	0.677	3.096	3.098	12.740	12.379
0.70	9	0.653	0.321	5.726	5.950	18.230	18.171
0.80	3	8.618	6.559	0.507	0.032	10.941	9.731
0.80	6	4.907	2.431	1.974	1.347	14.560	13.288
0.80	9	2.854	1.367	4.070	3.931	19.359	18.640

Table 5. 9: Hyper-Exponential Inter-Arrival Times and Hyper-Exponential Processing Times

ρ	Basestock Level	Average Backorders at Stage 3		Average Inventory at Stage 3		Average Intermediate Inventory	
		Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
0.60	3	3.171	1.505	1.095	0.826	8.171	7.142
0.60	6	1.176	0.509	3.463	3.592	12.848	12.380
0.60	9	0.473	0.219	6.277	6.523	18.311	18.159
0.70	3	7.224	4.598	0.638	0.157	10.828	9.187
0.70	6	3.818	1.625	2.406	2.186	14.647	13.195
0.70	9	1.925	0.842	4.787	5.001	19.293	18.586
0.80	3	16.972	14.196	0.265	0.001	17.386	15.353
0.80	6	10.817	6.944	1.265	0.312	19.486	16.789
0.80	9	6.682	3.569	2.983	2.222	22.703	20.506

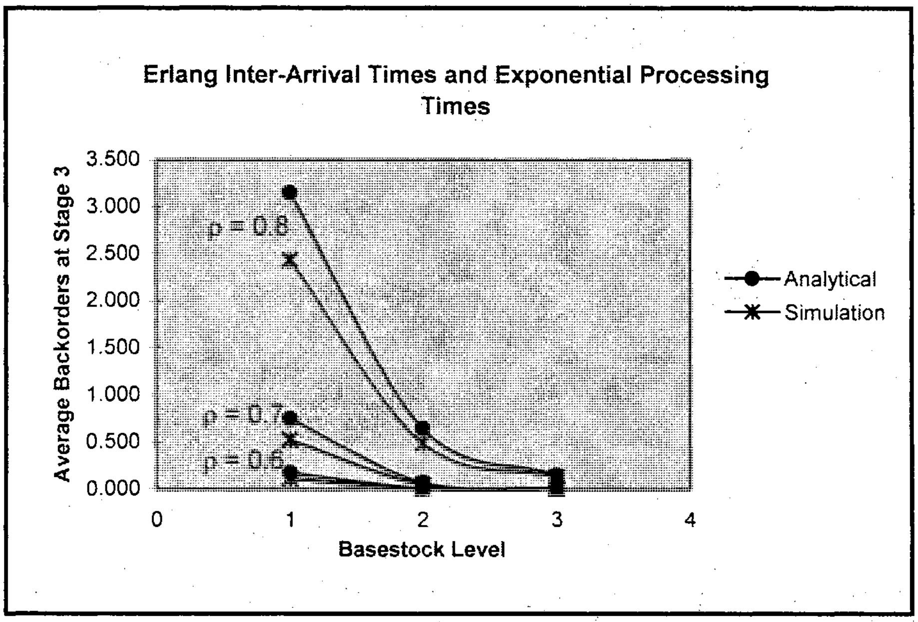


Figure 5. 5: E[B3] for Erlang Inter-Arrival Times and Exponential Processing Times

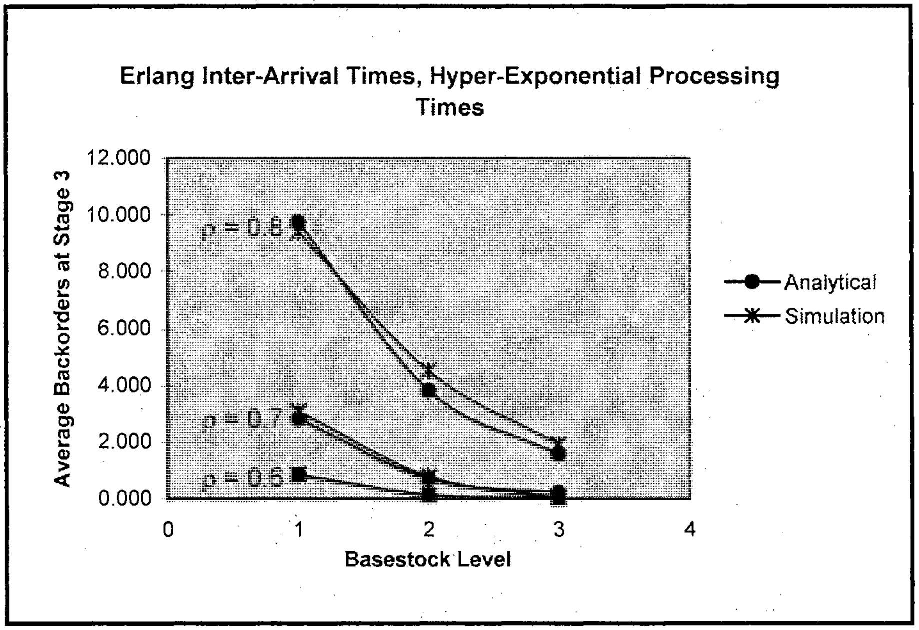


Figure 5. 6: E[B3] for Erlang Inter-Arrival times and Hyper-Exponential Processing Times

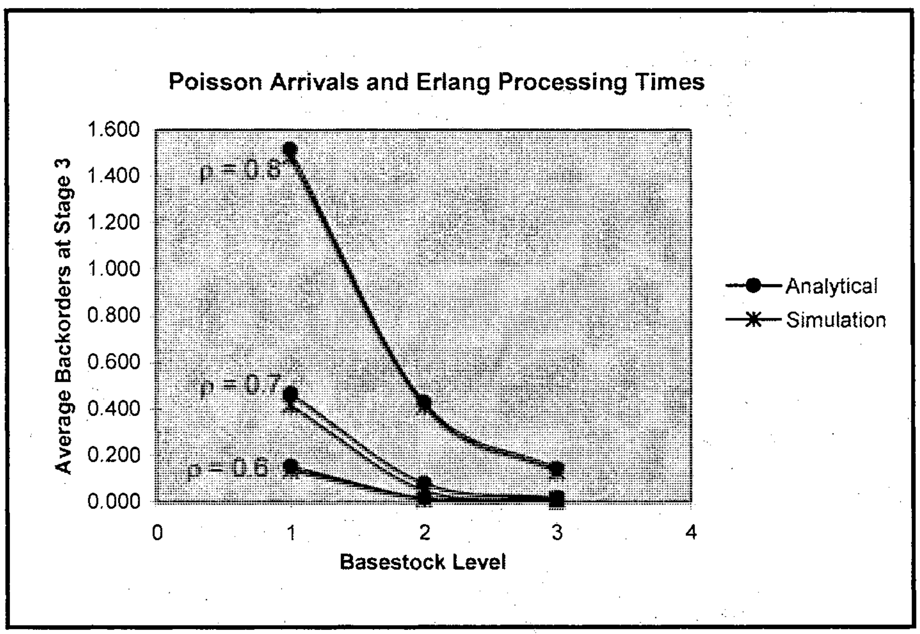


Figure 5. 7: E[B3] for Poisson Arrivals and Erlang Processing Times

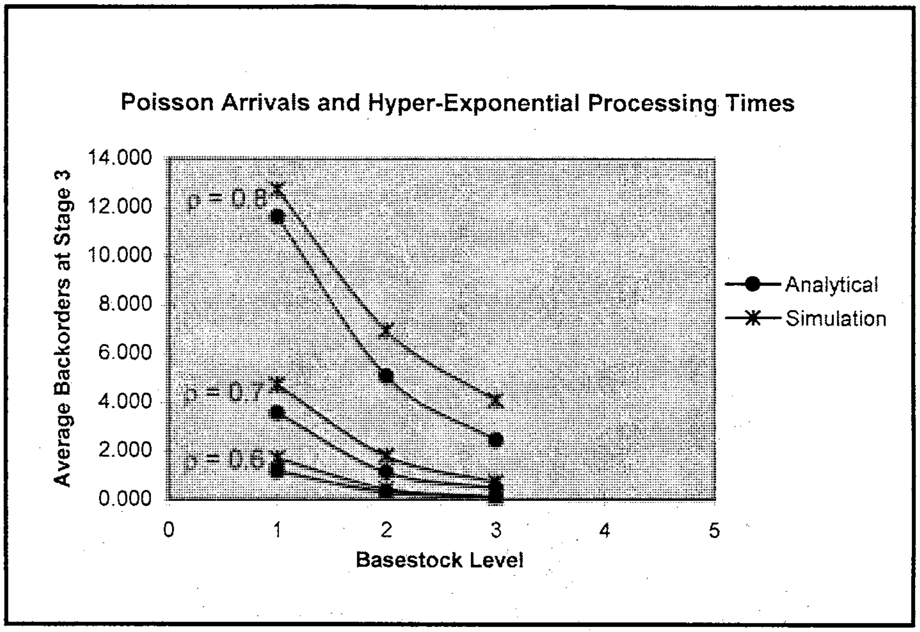


Figure 5. 8: E[B₃] for Poisson Arrivals and Hyper-Exponential Processing Times

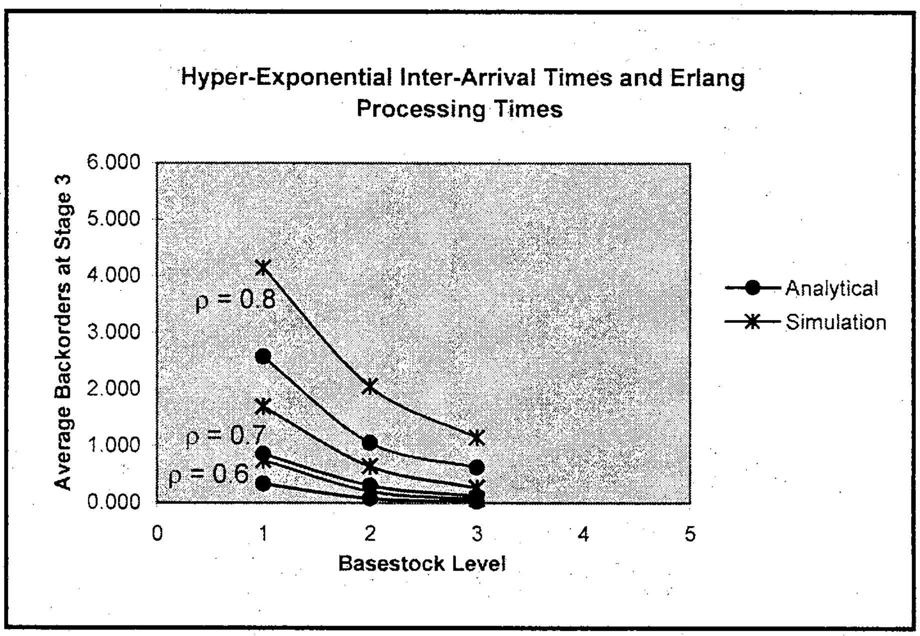


Figure 5. 9: E[B_3] for Hyper-Exponential Inter-Arrival Times and Erlang Processing Times

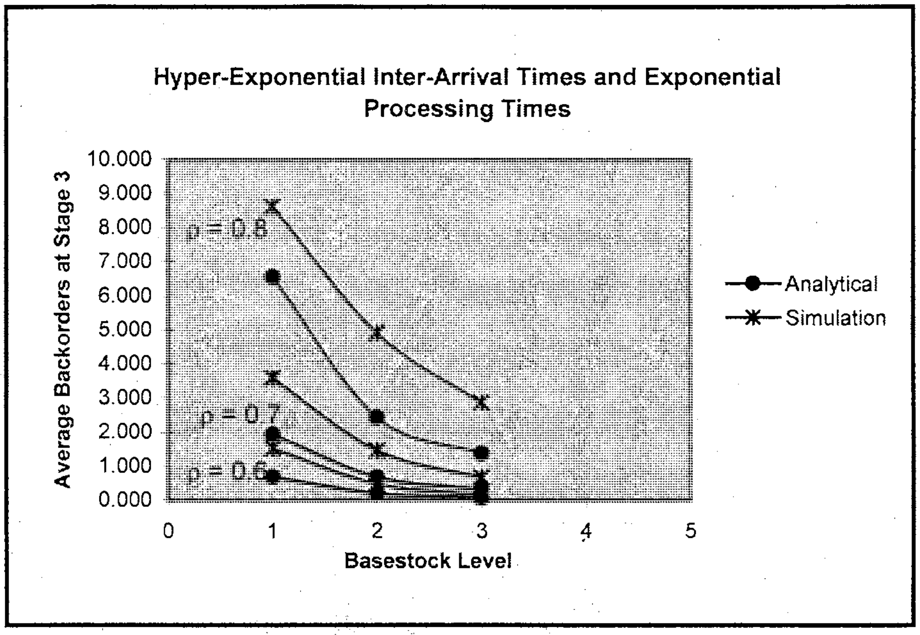


Figure 5. 10: E[B₃] for Hyper-Exponential Inter-Arrival Times and Exponential Processing Times

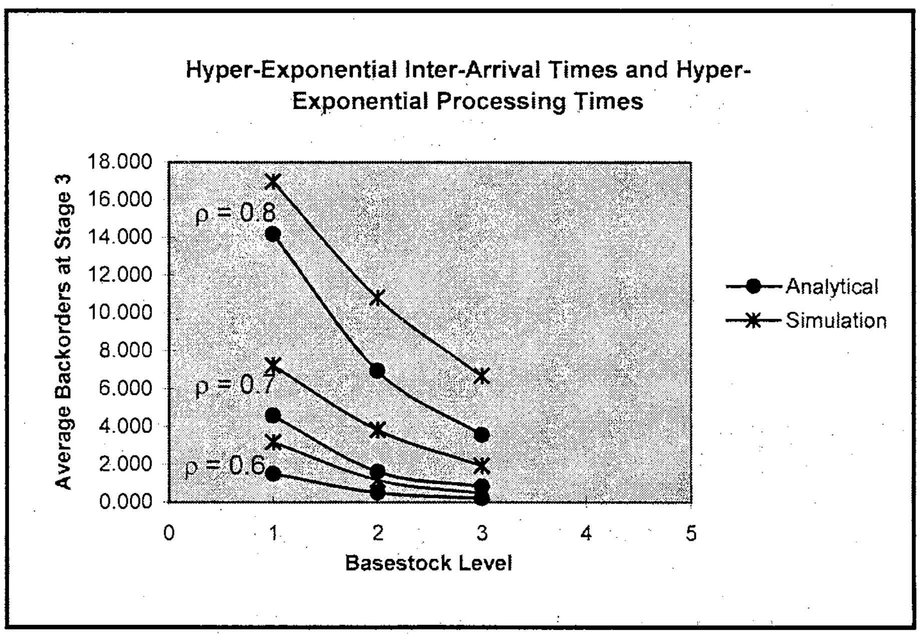


Figure 5. 11: E[B_3] for Hyper-Exponential Inter-Arrival Times and Hyper-Exponential Processing Times

Table 5. 10: Results for Ten-Stage Systems

				Average Backorders at Stage 10		Average Inventory at Stage 10	
Inter-Arrival Time SCV	Service Time SCV	Utilization	Basestock Level	Simulation	Analytical	Simulation	Analytical
2.25	0.25	0.7	3	1.814	0.807	1.225	1.047
2.25	0.25	0.7	6	0.642	0.310	3.543	3.675
2.25	1.00	0.7	3	6.215	1.844	0.729	0.541
2.25	1.00	0.7	6	1.597	0.676	3.040	3.116
2.25	1.00	0.8	3	5.097	1.814	0.748	0.451
2.25	1.00	0.8	6	2.258	1.075	2.583	2.412
2.25	1.00	0.8	3	0.166	1.043	1.566	0.773
0.25	1.00	0.8	3	7.792	10.265	0.509	0.001
0.25	2.25	0.8	3	40.502	43.024	0.001	0.000
1.00	0.25	0.8	3	1.659	1.444	1.025	0.585
1.00	1.00	0.8	3	13.026	11.634	0.241	0.000

5.4.2 Non-Homogeneous Systems

We examined six different configurations of non-homogeneous tandem systems. A particular combination of basestock level and SCV of processing time was used at each of the stages. The combinations were chosen such that the level of inventory was consistent with variability in the service process. That is, a node with higher (lower) SCV was provided with a larger (smaller) output store. The system tested was a three-stage system similar to the homogeneous systems presented in the earlier section. The arrival rate λ is set to unity so that desired utilizations of 0.70 and 0.80 are obtained by just modifying the mean processing times. The various configurations of service distribution SCV and basestock levels used at each stage are provided in Table 5.11. Three different inter-arrival time distributions namely Erlang, exponential and hyper-exponential were tested for each of the six configurations.

Table 5. 11: Configurations for the Non-Homogeneous Cases

Configuration	Stage 1		Stage 2		Stage 3	
	Service distribution SCV	Basestock level	Service distribution SCV	Basestock level	Service distribution SCV	Basestock level
1	0.25	3	1.00	6	2.25	9
2	0.25	3	2.25	9	1.00	6
3	1.00	6	0.25	3	2.25	9
4	1.00	6	2.25	9	0.25	3
5	2.25	9	0.25	3	1.00	6
6	2.25	9	1.00	6	0.25	3

Tables 5.12 through 5.14 present both simulation estimates and the analytical results for the various combinations of arrival distributions and utilizations. The results again indicate that the analytical model performs quite well (RPD < 11%) in most cases.

Table 5. 12: Non-Homogeneous Systems: Erlang Inter-Arrival Times

		Average Backorders at Stage 3		Average Inventory at Stage 3		Average Intermediate Inventory	
Configuration	ρ	Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
1	0.70	0.164	0.185	6.534	6.419	10.623	10.770
2	0.70	0.099	0.081	4.342	4.191	12.754	12.893
3	0.70	0.189	0.198	6.537	6.404	10.106	10.196
4	0.70	0.143	0.070	1.917	1.798	14.658	14.674
5	0.70	0.081	0.090	4.385	4.257	11.201	11.210
6	0.70	0.082	0.053	1.947	1.899	13.597	13.540
1	0.80	0.954	1.222	5.016	4.583	12.483	13.176
2	0.80	1.055	0.843	3.096	2.511	14.503	14.868
3	0.80	0.965	1.294	5.045	4.611	11.362	12.027
4	0.80	1.224	0.921	1.297	0.770	15.311	15.496
5	0.80	0.976	0.846	3.088	2.732	11.269	11.461
6	0.80	1.148	0.738	1.310	0.960	13.164	13.125

Table 5. 13: Non-Homogeneous Systems: Poisson Arrivals

		Average Backorders at Stage 3		Average Inventory at Stage 3		Average Intermediate Inventory	
Configuration	ρ	Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
1	0.70	0.473	0.443	5.930	5.777	10.818	10.945
2	0.70	0.534	0.334	3.689	3.554	13.182	13.059
3	0.70	0.501	0.446	5.821	5.685	10.364	10.428
4	0.70	0.713	0.489	1.482	1.324	14.879	14.832
5	0.70	0.497	0.322	3.677	3.542	11.476	11.426
6	0.70	0.653	0.450	1.500	1.400	13.770	13.696
1	0.80	2.099	2.005	4.167	3.752	13.114	13.454
2	0.80	2.687	1.740	2.434	1.799	15.458	15.141
3	0.80	2.376	1.894	3.985	3.630	12.335	12.265
4	0.80	3.205	2.243	0.900	1.072	16.343	15.845
5	0.80	2.587	1.431	2.372	1.801	12.277	11.629
6	0.80	3.197	1.937	0.890	0.487	14.130	13.450

Table 5. 14: Non-Homogeneous Systems: Hyper-Exponential Inter-Arrival Times

		Average Backorders at Stage 3		Average Inventory at Stage 3		Average Intermediate Inventory	
Configuration	ρ	Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
1	0.70	1.502	0.764	5.004	5.216	11.417	11.029
2	0.70	1.714	0.733	3.033	3.050	13.637	13.163
3	0.70	1.456	0.674	4.950	5.023	10.844	10.452
4	0.70	2.284	1.090	1.092	1.000	15.423	14.890
5	0.70	1.844	0.589	2.923	2.883	12.154	11.451
6	0.70	2.140	1.002	1.208	1.029	14.307	13.718
1	0.80	4.755	2.735	3.348	2.969	14.306	13.262
2	0.80	5.819	2.840	1.828	1.234	16.881	15.102
3	0.80	5.441	2.517	3.076	2.546	13.875	12.175
4	0.80	6.756	3.850	0.695	0.181	17.736	15.873
5	0.80	5.674	2.521	1.811	0.864	13.608	11.815
6	0.80	6.510	3.493	0.689	0.179	15.556	13.472

5.5 SUMMARY

This chapter extended the decomposition procedure used for the analysis of tandem make-to-stock systems with Poisson arrivals and exponential processing times to general arrivals and general processing times. The wide variety of example systems investigated show that the approximation works quite well in many situations. In the following chapter, it is shown that the same framework can be applied when additional manufacturing features are included as part of the system which alter the dynamics of system flow, with the focus still on tandem make-to-stock systems.

CHAPTER VI

MODELING TANDEM MAKE-TO-STOCK SYSTEMS: ADDITIONAL MANUFACTURING FEATURES

6.1 CHAPTER OVERVIEW

The previous chapter focused on generalizing the decomposition approach from a distributional perspective. This chapter focuses on extending the generalized model to include common manufacturing features such as batching of orders, multiple servers at a stage, limited supply of raw materials and service interruptions. First, the general procedure is outlined, and, then each subsequent section explains the modeling of a particular feature and how it can be included within the generalized procedure. In other words, a framework is established based on the approximation procedure developed in Chapters 4 and 5, and it is shown that this framework is versatile in that it can be applied to model many common manufacturing situations and features.

6.2 THE GENERALIZED PROCEDURE

The system under consideration is a tandem make-to-stock system where at each stage there is a stock of products at the output side. These products are subsequently used to make the product at the next stage unless it is the final stage where the product is the finished product, which is used to satisfy the external demand. The system is controlled by a basestock policy that is specified by the basestock level at each stage, which is the maximum planned inventory at the output side. The assumptions made for the system described in Chapter 5 hold true for all the systems described in this chapter as well. Demand occurs at the last stage and triggers the demand for the rest of the stages as described earlier. Demand inter-arrival times and processing times are stochastic, and are characterized by the first two moments, the mean and the squared coefficient of variation.

The solution procedure begins at Stage 1. For all systems except when there is a need to model the supply of raw materials, we solve for the approximate distribution of number of orders at this stage. The approximate distribution of inventory as well as the approximate distribution of backorders is then obtained from the distribution of number of orders. The procedure to determine the distribution of the number of orders varies depending on the manufacturing feature being modeled. The procedures for the different features are described in the following sections. First, the expected inventory level, the expected number of backorders and the probability that an order will not find a part at Stage 1 are obtained. From the second stage onwards, the delay model is used for the analysis. The procedure used to calculate the required performance measures at each stage is similar to the analysis of the delay model described in earlier chapters. The method used for determining the distribution of orders in the processing node changes with the manufacturing feature being modeled. The procedure to determine the arrival rate and squared coefficient of variation of the inter-arrival times to the processing node is also modified depending on the specific manufacturing feature being modeled. In short, as a new manufacturing feature is included, the procedure to determine the distribution of number in system for that specified type of queue is substituted in the general procedure and the analysis is then carried out.

6.3 MULTIPLE SERVERS AT A STAGE

Multiple machines/servers at a stage is a feature common in many production systems. The analysis of a tandem make-to-stock system with multiple-server stages follows the general procedure described in Section 6.2. In this section, the focus is on a procedure to find the distribution of number in system in a multiple-server queue. We use the procedure contained in Whitt's (1993) article on GI/G/m queues.

The following input parameters are required for the analysis of a GI/G/m queue.

λ, c_a^2 - rate and SCV of the inter-arrival time distribution;

τ, c_s^2 - mean and SCV of processing time distribution; and

m - number of servers at the node.

An approximation developed by Whitt (1993) for the probability mass function $P(N = n)$ where N is the number in system is given below:

$$P(N=n) = \begin{cases} P(Q = n-m) & n \geq m+1 \\ p(n) \big/ P(Q=0) & 0 \leq n \leq m \end{cases} \tag{6.1}$$

where Q is the queue length random variable, $p(n) = \dfrac{q(n)}{\sum_{j=0}^{m} q(j)}$ with $q(j) = \alpha^j e^{-\alpha} \big/ j!$. That is, $p(n)$ is a truncated Poisson distribution with intensity α. The intensity α is found out by matching the exact value of the expected number of busy servers:

$$\text{Expected number of busy servers} = m\rho = \sum_{n=0}^{m} n.P(N=n) + mP(Q>0) \tag{6.2}$$

leading to the formula

$$\sum_{n=0}^{m} np(n) = m[\rho - P(Q > 0)] \tag{6.3}$$

The parameter α is found out using the computational procedure developed by Jagerman (1984). The probability mass function $P(Q = n)$ is obtained by

$$P(Q = n) = P(Q > 0).\ P(C = n) \tag{6.4}$$

where C is the conditional queue length given that the queue is not empty. The procedure described in Whitt (1993) was used to determine $P(C = n)$ and $P(Q > 0)$.

The SCV of the departure process from a multiple server node is obtained using the standard approximation from QNA (Whitt, 1983).

$$c_d^2 \cong 1 + (1-\rho^2)(c_a^2 - 1) + \frac{\rho^2}{\sqrt{m}}(c_s^2 - 1) \tag{6.5}$$

In summary, the overall procedure begins with the calculation of $P(Q>0)$ and $P(C = n)$. It is then followed by the calculation of $p(n)$, $q(n)$ and α using the procedures described in Whitt (1993). Then, the equation (6.1) is used to find $P(N = n)$. Using $P(N = n)$, the distribution of inventory level and the number of backorders is obtained for a single-stage make-to-stock system with multiple servers. $P(N = n)$ is also used for determining the distribution of number in system in the delay model, which is then used in calculation of expected inventory level and expected number of backorders. This procedure is now used in conjunction with the general procedure outlined in Section 6.2 for the analysis of a tandem make-to-stock system with multiple servers at a stage.

6.3.1 Numerical Results

Thirty-six different configurations of tandem make-to-stock systems with multiple server stages were tested. Three inter-arrival distributions in combination with three service distributions, two levels of utilization, and two basestock levels were used to obtain the thirty-six different configurations. The mean demand arrival rate was set to one so that the mean service times could be manipulated to obtain the desired utilization. The SCV of the inter-arrival and service distributions used were 0.25 (Erlang with four stages), 1 (exponential), and 2.25 (hyper-exponential distribution), and the two utilizations used were 0.70 and 0.80. The number of servers at each stage was set to 3. The intermediate inventory measure includes everything except the orders in queue and being processed at the first stage and the finished goods. The results are presented in Tables 6.1 through 6.3. As before, RPD is used to evaluate the accuracy of the analytical model.

$$\text{Relative percent difference (RPD)} = \frac{\text{Analytical Result - Simulation Estimate}}{\text{Simulation Estimate}} * 100$$

In terms of estimating the average intermediate inventory, the model performed extremely well (less than 5% RPD) in 18 cases examined, very well (5% to 10% RPD) in 7 cases, and reasonably well (10% to 15% RPD) in 6 cases. In 35 of the 36 different configurations examined, the RPD for average intermediate inventory was less than 20%. In terms of number of backorders and inventory level, the RPD measure is sometimes inappropriate as the values involved are extremely small (Whitt, 1983). In case of systems where there is reasonably large backorder or inventory level, the model estimates the average values quite accurately. Overall, the model tends to capture the behavior of the system as a function of the arrival and service parameters and the basestock levels.

Table 6. 1: Multiple-Server System with Erlang Inter-Arrival Times

System Parameters			Average Backorders at Stage 3		Average Inventory at Stage 3		Average Intermediate Inventory	
C_s^2	ρ	Basestock Level	Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
0.25	0.70	3	0.316	1.276	0.775	0.687	6.259	7.271
0.25	0.70	6	0.003	0.346	3.720	3.610	11.994	12.418
0.25	0.80	3	1.153	3.103	0.382	0.303	6.863	8.853
0.25	0.80	6	0.043	0.869	3.108	2.816	12.031	13.106
1.00	0.70	3	1.341	2.143	0.584	0.506	7.118	7.964
1.00	0.70	6	0.122	0.821	3.356	3.232	12.127	12.699
1.00	0.80	3	4.314	5.628	0.206	0.104	9.287	10.713
1.00	0.80	6	0.897	1.820	2.256	1.969	12.879	14.039
2.25	0.70	3	3.578	3.765	0.422	0.255	8.878	9.189
2.25	0.70	6	0.917	1.219	2.718	2.573	12.918	13.398
2.25	0.80	3	10.124	10.407	0.119	0.011	13.831	14.066
2.25	0.80	6	4.770	4.641	1.346	0.883	16.080	16.429

Table 6. 2: Multiple-Server System with Poisson Arrivals

System Parameters			Average Backorders at Stage 3		Average Inventory at Stage 3		Average Intermediate Inventory	
C_s^2	ρ	Basestock Level	Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
0.25	0.70	3	1.255	1.919	0.718	0.502	6.703	7.579
0.25	0.70	6	0.158	0.605	3.278	3.173	12.042	12.593
0.25	0.80	3	3.184	4.443	0.374	0.134	7.753	9.264
0.25	0.80	6	0.742	1.519	2.415	2.085	12.280	13.389
1.00	0.70	3	2.674	2.900	0.563	0.333	7.841	8.319
1.00	0.70	6	0.640	0.948	2.897	2.787	12.460	12.913
1.00	0.80	3	7.099	7.222	0.232	0.037	10.826	11.196
1.00	0.80	6	2.646	2.840	1.808	1.373	13.899	14.478
2.25	0.70	3	5.234	4.452	0.430	0.169	9.901	9.451
2.25	0.70	6	1.844	1.629	2.418	2.264	13.524	13.534
2.25	0.80	3	13.128	11.944	0.145	0.004	15.449	14.524
2.25	0.80	6	6.945	5.734	1.208	0.579	17.296	16.740

Table 6. 3: Multiple-Server System with Hyper-Exponential Inter-Arrival Times

System Parameters			Average Backorders at Stage 3		Average Inventory at Stage 3		Average Intermediate Inventory	
C_s^2	ρ	Basestock Level	Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
0.25	0.70	3	2.956	2.781	0.707	0.306	7.261	7.880
0.25	0.70	6	1.012	1.180	2.771	2.650	12.190	12.935
0.25	0.80	3	6.472	6.430	0.390	0.035	8.771	9.682
0.25	0.80	6	2.888	2.968	1.760	1.036	12.672	14.218
1.00	0.70	3	4.867	3.869	0.556	0.186	8.897	8.651
1.00	0.70	6	1.929	1.600	2.473	2.251	13.060	13.317
1.00	0.80	3	10.580	9.468	0.257	0.008	12.211	11.767
1.00	0.80	6	5.833	4.789	1.474	0.599	15.137	15.497
2.25	0.70	3	7.654	5.587	0.441	0.083	11.073	9.819
2.25	0.70	6	3.631	2.503	2.130	1.641	14.501	14.178
2.25	0.80	3	17.713	14.488	0.173	0.001	18.040	15.223
2.25	0.80	6	10.504	8.740	1.038	0.138	19.720	18.377

6.4 BATCH SERVICE

In many manufacturing systems, workstations process parts in batches to reduce the effect of set-up times and to make efficient use of resources such as tools and operators. In this section, we show that this manufacturing feature can be modeled within the decomposition framework. The system under consideration here is limited to the tandem make-to-stock systems even though the procedure can be applied to other systems, such as feed-forward networks. The assumptions that are specific to batch service systems are summarized next.

The incoming orders are for single parts both at the final stage as well as at the intermediate stages. The orders at an intermediate stage consume individual parts from the previous stage, and then proceed to be batched before being processed at that stage. The parts after being processed as a batch are immediately split into individual parts before proceeding to the output store of the stage.

The procedure begins at Stage 1 where the arrival rate and SCV of the inter-arrival distribution are known. The processing time distribution for a batch is also known. The distribution of the number in system at Stage 1 is computed from which the distribution of backorders and inventory levels are calculated. It is reminded that the stage processes orders in batches. The procedure to find the distribution of the number in system in a batch node is presented in the next sub-section. The departure rate from this stage is the same as the external arrival rate and the SCV of the inter-departure time distribution is computed from the service and arrival parameters using the procedure in Whitt (1983). Also, the probability that an order is backordered is calculated. Beginning at Stage 2 till the last stage and proceeding sequentially, the delay model is repeatedly used to obtain the distribution of backorders and that of the inventory level. The distribution of the number in system at the processing node in the delay model is now the distribution of the number in system in a batch processing node.

6.4.1 The Distribution of the Number in System in a Batch Processing Node

This section describes the approximate procedure to determine the distribution of the number in system in a single stage batch processing make-to-stock system. This procedure is an extension of the procedure to estimate the mean number of jobs in ordinary batch service queues described by Bitran and Tirupati [1989]. The method first explains the derivation of the expression for mean number of jobs which is then followed by the expression for the distribution of the number in the system.

The single-server make-to-stock system with batch processing is shown in Figure 6.1. Station 0 is a fictitious station that can be interpreted as a staging area for forming batches. Demand arriving to the system is satisfied from the output store if parts are available else it is backordered. The basestock policy identified by the basestock level S is one for one and thus, each demand arrival triggers an order to replenish the consumed part. The orders pick up raw parts which are assumed to be always available. The orders then proceed to Station 0 where they wait until a batch of size r is formed. Accumulation of r orders signifies the completion of a batch which is transferred immediately to the queue at Station 1.

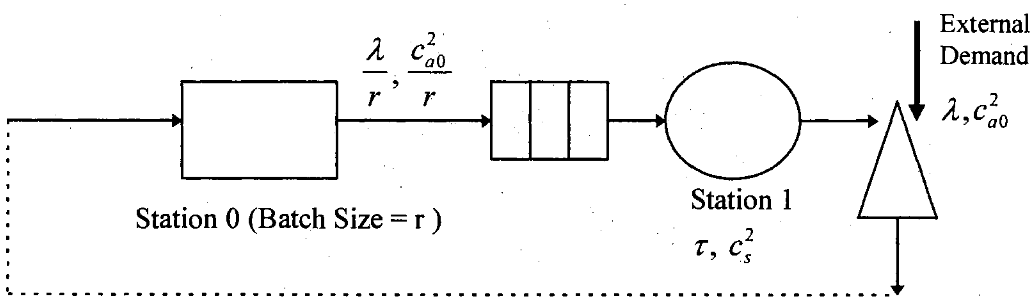


Figure 6. 1: A Single-Server Batch Processing Make-to-Stock System

The following notation is used in the estimation the mean number of orders:

λ - demand arrival rate;

c_{a0}^2 - SCV of the inter-arrival time distribution;

τ - mean processing time;

c_s^2 - SCV of the processing time distribution;

N - number of orders in the system at any time;

N_0 - number of orders at Station 0 (batching station) at any time; and

N_1 - number of batches at Station 1 at any time.

We assume that the external arrival process is a renewal process. This implies that the arrivals at Station 0 follow a renewal process which yields the following results. $E[N_0]$, the mean number of orders or parts at Station 0 is $(r-1)/2$ (Bitran and Tirupati, 1988). The arrival rate to Station 1 is given by λ/r and the SCV of inter-arrival times at Station 1 is equal to c_{a0}^2/r (Bitran and Tirupati, 1988).

Station 1 is now modeled as an ordinary GI/G/1 queue where each customer represents a batch of r orders. Using the approximation given in Chapter 5, $E[N_1]$, the mean number of batches in Station 1, is determined.

Now, $E[N]$, which is the average number of orders in the system is given by $E[N] = E[N_0] + E[N_1].r$. The utilization of Station 1, ρ, is equal to $(\lambda/r).\tau$.

Define $\sigma = (E[N] - \rho)/E[N]$

The approximate probability distribution of N is given by (Buzacott and Shanthikumar, 1993),

$$P(N = n) = \begin{cases} 1-\rho, & n=0 \\ \rho(1-\sigma)\sigma^{n-1}, & n = 1, 2, \ldots \end{cases} \tag{6.6}$$

From $P(N = n)$, the distribution of backorders and inventory can be easily derived as shown in Chapter 5.

The departure process from the batch node needs to be computed so that the arrival process to the next stage can be obtained. At Station 1, we know the first two moments of the inter-arrival and service times. Using the procedure described by Whitt (1983), the parameters of the departure process are computed. The mean rate of the departure process is same as the arrival process and thus after the splitting of batches, the mean departure rate of orders remains as λ. Denote the SCV of the departure process before splitting by c_{db}^2. The departure process after the splitting into individual parts is simply $c_d^2 = c_{db}^2.r$ (Whitt, 1983). Using c_d^2, the effective inter-arrival time SCV at the next stage is computed using the procedure described in Chapter 5.

We now know all of the required procedures to determine the distribution of the number in system and hence, the distribution of the inventory level and the number of backorders in a batch processing make-to-stock system. This procedure is used within the framework established in Section 6.2 to complete the analysis of a tandem make-to-stock system with batch processing at the individual stages.

6.4.2 Numerical Results

The parameters of the six configurations used to test the batch service case are presented in Table 6.4. The simulation estimates were obtained by averaging estimates from ten replications. Each replication had a run-length of 50,000 time units in addition to 5,000 time units of warm-up. Three different inter-arrival time distributions were used in combination with the six configurations giving a total of 18 different test cases. The three inter-arrival distributions used were Erlang with SCV = 0.25, exponential, and hyper-exponential with SCV = 2.25. The batch size was set at 5 at all the stages and across all the configurations. The analytical and simulation results for the 18 test cases are presented in Tables 6.5 through 6.7.

The results indicate that the approximation performs very well (less than 10% RPD) for 12 of the 18 cases examined in terms of expected number of backorders. In terms of average intermediate inventory, the model performs extremely well (less than 5% RPD) in 13 of the 18 cases examined. In terms of average inventory at stage 3, the percentage difference is not a good measure since the numbers are relatively small in value. Overall, the approximation for the tandem make-to-stock system with batch processing performs accurately.

Table 6. 4: Configurations for Testing the Batch Processing Extension

Configuration	Stage 1 Service Distribution SCV	Stage 1 Basestock Level	Stage 2 Service Distribution SCV	Stage 2 Basestock Level	Stage 3 Service Distribution SCV	Stage 3 Basestock Level
1	0.25	9	1.00	12	2.25	15
2	0.25	9	2.25	15	1.00	12
3	1.00	12	0.25	9	2.25	15
4	1.00	12	2.25	15	0.25	9
5	2.25	15	0.25	9	1.00	12
6	2.25	15	1.00	12	0.25	9

Batch Size = 5 at all the stages

Table 6. 5: Batch Processing System with Erlang Inter-Arrival Times

	Average Backorders at Stage 3		Average Inventory at Stage 3		Average Intermediate Inventory	
Configuration	Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
1	18.327	20.682	2.438	1.922	38.486	47.306
2	21.670	22.108	1.274	0.071	47.343	50.583
3	19.788	19.345	1.952	2.262	35.235	39.804
4	23.716	23.311	0.526	0.001	44.368	46.031
5	25.782	20.148	0.821	0.066	34.758	32.847
6	26.339	23.909	0.406	0.001	34.869	36.673

Table 6. 6: Batch Processing System with Poisson Arrivals

	Average Backorders at Stage 3		Average Inventory at Stage 3		Average Intermediate Inventory	
Configuration	Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
1	21.690	22.205	2.185	1.522	43.039	47.842
2	25.081	24.113	1.220	0.038	48.582	51.234
3	22.635	20.793	1.785	1.701	36.180	40.308
4	26.500	25.603	0.527	0.001	45.755	46.819
5	28.340	22.420	0.793	0.024	34.478	33.630
6	28.181	26.392	0.439	0.000	36.852	37.626

Table 6. 7: Batch Processing System with Hyper-Exponential Inter-Arrival Times

	Average Backorders at Stage 3		Average Inventory at Stage 3		Average Intermediate Inventory	
Configuration	Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
1	25.961	24.628	2.013	0.977	46.764	48.440
2	30.446	27.265	1.128	0.013	52.045	52.041
3	28.038	23.305	1.624	0.936	39.603	41.166
4	32.088	29.300	0.524	0.000	48.425	48.096
5	33.773	26.031	0.811	0.004	37.556	34.826
6	32.836	30.177	0.467	0.000	38.043	38.976

6.5 MAKE-TO-STOCK SYSTEMS WITH A LIMITED SUPPLY OF RAW MATERIALS

In all the systems considered thus far, it was assumed that raw materials were always available at Stage 1. In this section, this assumption is relaxed and a limited supply of raw materials is modeled at the input side of Stage 1. The orders that arrive at Stage 1 now pick up raw materials from a raw material store with a limited capacity. We assume that the one-for-one replenishment policy is also followed for the raw materials. The replenishment time is random with a mean of γ. Also, let S_0 represent the maximum amount of raw material stock. The rest of the dynamics is similar to the tandem make-to-stock systems described in Chapters 4 and 5.

The analysis of the system follows the procedure described in Chapter 5 except for the analysis pertaining to Stage 1. Stage 1 is now modeled using the delay model. The supply process is modeled by an infinite server (delay) node, and using M/G/∞ formulas, the expected number of backorders, the expected inventory level and the probability of an order being backordered are determined.

We require the expected number of backorders at the raw material store in the analysis of Stage 1. The probability p of an order not finding raw materials is assumed to be the steady-state probability of the raw material store being empty. This is equivalent to having S_0 or more busy servers in the M/G/∞ system. Hence, we have

$$p = 1 - \sum_{k=0}^{S_0-1} \frac{(\lambda\gamma)^k . e^{-\lambda\gamma}}{k!} \tag{6.7}$$

The expected raw material inventory level is given by

$$E[I_0] = \sum_{k=0}^{S_0-1} (S_0 - k). \frac{(\lambda\gamma)^k . e^{\lambda\gamma}}{k!} \tag{6.8}$$

The expected number of backorders is given by

$$E[B_0] = (\lambda\gamma) + E[I_0] - S_0 \tag{6.9}$$

The expected number of backorders thus calculated becomes the average number (ρ_d) in the delay node of the delay model described in Chapters 4 and 5. Using the procedure developed for the delay model, performance measures for Stage 1 are computed. Beginning at Stage 2, the generalized approximation procedure is now used for the analysis of the entire system.

6.5.1 Numerical Results

We use homogeneous configurations to test the approximations for systems with limited raw materials supply. The lead time at the supplier is random with a mean of 3 time units. The demand arrival process is Poisson with a mean of one. Three different service processes were used in combination with two levels of raw material basestock levels and three different basestock levels for intermediate and finished parts, resulting in a total of eighteen different combinations. Comparisons of the analytical results with simulation estimates are presented in Table 6.8.

Table 6. 8: Results for Make-to-Stock Systems with a Limited Supply of Raw Materials

	Basestock Level		Average Backorder at Stage 3		Average Inventory at Stage 3		Average Intermediate Inventory	
c_s^2	S_0	S_i	Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
0.25	1	1	5.707	5.573	0.075	0.001	3.881	3.722
0.25	1	3	1.804	1.602	0.954	0.445	5.071	5.308
0.25	1	6	0.428	0.434	3.448	3.190	10.127	10.394
0.25	3	1	4.301	4.483	0.086	0.005	3.981	4.006
0.25	3	3	1.472	1.527	1.050	0.553	6.201	6.502
0.25	3	6	0.396	0.431	3.492	3.218	11.702	11.740
1	1	1	10.990	11.077	0.017	0.000	7.973	8.027
1	1	3	6.021	5.586	0.458	0.061	8.532	8.475
1	1	6	2.154	1.574	2.470	1.950	11.695	11.574
1	3	1	9.295	9.779	0.034	0.000	7.817	8.107
1	3	3	5.080	4.800	0.562	0.112	9.067	9.016
1	3	6	1.990	1.499	2.539	2.069	13.002	12.758
2.25	1	1	19.663	19.785	0.007	0.000	14.767	14.735
2.25	1	3	14.109	13.571	0.185	0.001	14.894	14.520
2.25	1	6	7.854	5.771	1.335	0.507	16.435	15.214
2.25	3	1	18.063	18.254	0.013	0.000	14.649	14.582
2.25	3	3	13.097	12.180	0.245	0.004	15.218	14.504
2.25	3	6	7.076	5.250	1.420	0.644	17.127	15.934

The results indicate the model accurately estimates the average intermediate inventory in every case tested. The RPD was less than 10% in all the cases. In terms of expected backorders, the model performs very well (less than 10% RPD) in 13 of the 18 cases examined. Overall, the analytical model compares very well with simulation thus showing that the framework can be easily extended for modeling make-to-stock systems with a limited supply of raw materials.

6.6 MULTIPLE-CLASS MAKE-TO-STOCK SYSTEMS

In this section, analytical models for make-to-stock systems with multiple classes of customers are developed. The system under consideration has K classes of customers with λ_k and c_{0k}^2 being the parameters of the class k demand arrival process, $k = 1, 2, \ldots, K$. The focus is again limited to tandem systems even though the procedure can be easily extended to other production networks such as feed-forward networks. At each stage in the system, the service time distribution is unique to each class and is described by the mean processing time, τ_{jk}, and SCV, c_{skj}^2, for class k, $k = 1, 2, \ldots, K$, and stage j, $j = 1, 2, \ldots, M$.

The M-stage tandem system is similar to the system described in Chapter 4. The basestock policy is specific to a particular class and is represented by a non-negative integer, S_{jk}, where, $j = 1, 2, \ldots, M$, and $k = 1, 2, \ldots K$. Additionally, it is assumed that $S_{j1} = S_{j2} = \ldots = S_{jK}$. That is, the basestock level is the same for all the classes at a stage. Demand for a particular class k is satisfied at stage M from the available stock. If there is no stock of class k type units at the last stage, the demand is backordered. Production at a stage continues until the inventory levels of S_{jk} for all classes k is reached. The orders for a class j at stage M looks into the stage M-1 store for class j parts. If parts are available, the order picks the part and joins the queue at stage M for processing. Otherwise, the request for part is backordered, and the order waits for an order in that class to finish processing at stage M-1.

At stage 1, orders immediately join the queue as it is assumed that raw materials that are required for each class are always available.

The approach used in QNA (Whitt, 1983) to handle multiple customer classes is modified for use in a make-to-stock system. The analysis of the system begins by aggregating the arrival distributions, processing distributions and basestock levels into an equivalent single class system. The procedure described in Chapter 5 is then used to analyze this aggregated single class system. After the analysis, a disaggregation procedure is used to compute the detailed performance measures. Only the aggregation and disaggregation procedures are described here.

The Aggregation Procedure

The basic aggregation procedure is a modification of the procedure originally developed by Whitt (1983) for ordinary queueing networks. The following notation is used:

λ_k, c_{0k}^2	rate and SCV of the inter-arrival distribution for class *k;*
τ_{jk}, c_{sjk}^2	mean and SCV of the processing time distribution for class *k* at stage *j;*
S_{jk}	basestock level for class *k* at stage *j;*
λ, c_0^2	rate and SCV of the aggregated arrival distribution;
τ_j, c_{sj}^2	mean and SCV of the aggregated processing time distribution at stage *j;*
S_j	aggregated basestock level at stage *j;*

B_{jk} average number of backorders for class k at stage j;

N_{jk} average number of orders for class k at stage k;

B_j aggregated average number of backorders at stage k; and

N_j aggregated average number of orders at stage k.

In the above quantities, $k = 1, 2, \ldots, K$ and $j = 1, 2, \ldots, M$.

The aggregated arrival rate and SCV are given by

$$\lambda = \sum_{k=1}^{K} \lambda_k \text{ and } c_{0a}^2 = \frac{\sum_{k=1}^{K} c_{0k}^2 \lambda_k}{\sum_{k=1}^{K} \lambda_k} \tag{6.10}$$

The aggregated mean processing time and SCV at stage j are given by

$$\tau_j = \frac{\sum_{k=1}^{K} \tau_{jk} \lambda_k}{\sum_{k=1}^{K} \lambda_k}; \; \tau_j^2 (c_{sj}^2 + 1) = \frac{\sum_{k=1}^{K} \tau_{jk} (c_{sjk}^2 + 1) \lambda_k}{\sum_{k=1}^{K} \lambda_k} \tag{6.11}$$

The aggregated basestock level at each stage is given by $S_j = \sum_{k=1}^{K} S_{jk}$. This completes the aggregation procedure.

The Disaggregation procedure

The performance measures obtained using the approximation procedure in Chapter 5 are for an aggregated single class model. The performance measures of interest are the expected number of backorders and the expected inventory levels at each stage. These measures are obtained as follows:

Let $E[B_j]$ and $E[N_j]$ be the average number of backorders and the average number of orders at stage j, respectively. These are obtained from the aggregated single class model. The class specific measures are

$$E[B_{jk}] = E[B_j]\frac{\lambda_k}{\sum_{k=1}^{K}\lambda_k}; \text{ and } E[N_{jk}] = (E[W_j] + \tau_{jk}).\lambda_k + E[B_{j-1k}] \quad (6.12)$$

$$E[I_{jk}] = S_k + E[B_{jk}] - E[N_{jk}] \quad (6.13)$$

where $E[W_j]$ is the average time in queue at stage j.

This completes the disaggregation procedure.

In summary, the analysis of a multiple-part type make-to-stock system begins with the aggregation procedure. It is then followed by the analysis of the aggregated single class system using the procedure described in Chapter 5. The final step is the disaggregation procedure which gives the required class-specific performance measures.

6.6.1 Numerical Results

Three-stage systems with two part types were used to test the approximations developed. The arrival rates of the two part types were chosen to be 1.036 for part type 1 and 0.964 for part type 2 so that the average of the arrival rates was equal to one. The processing time distribution for a part type was chosen to be the same at all the stages in the system. The basestock levels were also chosen to be the same for each part type across all the stages. The basestock level was chosen to be 3 at all the stages across all part types. The mean processing times for each part type at each stage were chosen to be 0.77 for part type 1 and 0.83 for part type 2 so that the aggregated mean processing times gave the desired utilization of 0.80 at all the stages. In total, ten different configurations were tested and the results comparing the expected number of backorders and the average inventory level of each part type obtained using the simulation and analytical models are presented in Table 6.9 and 6.10.

The results from the analytical models are comparable to the simulation models. The average inventory levels were accurately determined by the analytical model in 5 of the 10 cases examined. These results indicate the approximation scheme is a good rough-cut tool for the evaluation of multiple-part type make-to-stock systems.

Table 6. 9: Results for Multiple-Class Make-to-Stock Systems

Inter-Arrival Time SCV		Processing Time SCV		Average Backorders at Stage 3 - Part Type 1		Average Inventory at Stage 3 - Part Type 1	
c_{01}^2	c_{02}^2	c_{si1}^2	c_{si2}^2	Simulation	Analytical	Simulation	Analytical
0.25	0.25	0.25	0.25	0.002	0.007	2.367	2.400
0.25	0.25	0.25	1	0.165	0.098	1.975	2.127
0.25	0.25	1	0.25	0.029	0.063	2.171	2.184
0.25	0.25	1	1	0.368	0.258	1.661	2.002
0.25	1	0.25	0.25	0.029	0.094	2.171	2.135
0.25	1	1	1	0.368	0.471	1.661	1.973
1	1	0.25	0.25	0.079	0.172	2.130	2.060
1	1	0.25	1	0.489	0.375	1.682	1.990
1	1	1	0.25	0.172	0.317	1.864	1.982
1	1	1	1	0.740	0.591	1.535	1.975

Table 6. 10: Results for Multiple-Class Make-to-Stock Systems (Continued)

Inter-Arrival Time SCV		Processing Time SCV		Average Backorders at Stage 3 - Part Type 2		Average Inventory at Stage 3 - Part Type 2	
c_{01}^2	c_{02}^2	c_{si1}^2	c_{si2}^2	Simulation	Analytical	Simulation	Analytical
0.25	0.25	0.25	0.25	0.020	0.010	2.080	2.100
0.25	0.25	0.25	1	0.544	0.147	1.572	1.690
0.25	0.25	1	0.25	0.337	0.094	1.718	1.776
0.25	0.25	1	1	1.561	0.387	1.185	1.502
0.25	1	0.25	0.25	0.337	0.141	1.718	1.703
0.25	1	1	1	1.561	0.706	1.185	1.459
1	1	0.25	0.25	0.108	0.258	1.863	1.591
1	1	0.25	1	1.002	0.562	1.267	1.485
1	1	1	0.25	0.628	0.512	1.412	1.500
1	1	1	1	1.932	0.887	1.094	1.485

6.7 MODELING SERVICE INTERRUPTIONS

In Chapter 3, it was mentioned that failures can be classified into two types. Failures that last for a short time such as service interruptions can be modeled within the performance model. Failures that last for a long time such as major equipment failures are better handled using techniques such as performability analysis. Service interruptions can be further classified into interruptions caused by machine and by parts. Examples of machine interruptions include planned maintenance and minor faults. Part interruptions include those due to jamming of parts and tool breakage. Each type of interruption is modeled differently as shown below. It is common practice to model service interruptions by approximately capturing their effect on system performance through the modification of service times. We describe these modifications using the notation of Segal and Whitt (1989) and Suri et al. (1993).

Machine specific interruptions: The availability of a machine is modeled by an alternating renewal process; that is, there is a succession of intervals U_1, D_1, U_2, D_2,... during which the machine is alternating up (available for service) and down (unavailable for service). It is assumed that these up and down times are mutually independent, for all machines as well as within each sequence for each machine. The down times are characterized by their mean and squared coefficient of variation and the up times are characterized by their mean.

This model is analyzed approximately as if the down times are triggered by processing times. As mentioned earlier we modify the processing times using the procedure suggested by Segal and Whitt (1989). Each product causes a down time with probability p, and so has a expanded processing time equal to the original processing time plus an independent down time, and has an ordinary processing time with probability (1-*p*). The modified model is now a standard GI/G/1 queue and can be handled using the two-moment approximations. The down time probability p is first chosen in order to produce the proper traffic intensity, and subsequently to capture the principal effect of the increased variability a revised processing time variability parameter is calculated. The adjusted processing time variability parameter also affects the approximation for the departure process, and thus, the other stages in the network. The modified processing time distribution parameters $\overline{\tau}$ and $\overline{c_s^2}$ are given by

$$\overline{\tau} = \tau + pd \tag{6.14}$$

and $$\overline{\tau}^2(\overline{c_s^2} + 1) = (c_s^2 + 1)\tau^2 + p[c_d^2 d^2 + 2d\tau + d^2] \tag{6.15}$$

where d is the mean and c_d^2 is the SCV of the down time. The down time probability is chosen so that the new traffic intensity is appropriately related to the original traffic intensity by the relationship $\overline{\rho} = \rho + d/(d+u)$ where u is mean up time. That is, the proportion of time that the machine is busy now includes the proportion of time that the machine is down. Now we have, $\overline{\tau} = \overline{\rho}/\lambda$. Using (6.14) and (6.15), $\overline{c}_s^2$ is obtained. $\overline{\tau}$ and $\overline{c}_s^2$ replace the original processing time distribution parameters for each of the stages and the analysis procedure described in earlier chapters is carried out.

<u>Part specific interruptions</u>: In situations where interruptions are due to parts, it is assumed that during the service of each part, there is a fixed probability p that a part might

cause a random delay *d*. The distribution of delay is specified by the mean, *d* and SCV, c_d^2. Then, equations (6.14) and (6.15) are used to modify the processing time parameters.

6.7.1 Numerical Results

As discussed earlier, service interruptions can be classified into two types: interruptions that are caused by the machine and those that are caused by the products. Results are presented for both cases.

Tandem Systems with Machine-Related Interruptions: Six configurations of a three-stage homogeneous system were tested. The demand arrival rate was set to one in all examples. Two different inter-arrival time distributions, namely, Erlang distribution with 4 stages and the exponential distribution, along with three different basestock levels, S = 1, 3 and 6, were used to give a total of six combinations. The processing time parameters remained the same in all the configurations but were different across the stages. The processing time parameters including the mean up-time and downtime distribution are presented in Table 6.11. The processing time parameters are chosen so that the effective utilization is the same (= 0.70) for all the stages.

Table 6. 11: Machine Parameters for Systems with Machine-Related Interruptions

	Stage 1	**Stage 2**	**Stage 3**
Processing time distribution	Mean = 0.65 SCV = 0.25	Mean = 0.60 SCV = 0.25	Mean = 0.55 SCV = 0.25
Up time distribution	Uniform(90, 100)	Uniform(85, 95)	Uniform(80,90)
Down time distribution	Uniform(1, 9)	Uniform(5, 15)	Uniform(10, 20)

The results presented in Table 6.12 indicate that the analytical approximation did not perform as well as it did for other features. In terms of average intermediate inventory, the RPD was larger than 10% in the all the cases examined. A detailed investigation of the intermediate calculations in the simulation and analytical models indicated that the SCV of the modified service times were not captured accurately by the analytical model. This in turn affected the accuracy of the departure process SCV and hence, the arrival process SCV to every stage. Extensive numerical testing is needed before any modification to the approximation can be made. Such an investigation was beyond the scope of this effort because of time and resource constraints.

Table 6. 12: Results for Systems with Machine-Related Interruptions

System Parameters			Average Backorders at Stage 3		Average Inventory at Stage 3		Average Intermediate Inventory	
c_{0a}^2	c_{sj}^2	S	Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
0.25	0.25	1	4.838	7.877	0.146	0.014	6.551	9.404
0.25	0.25	3	2.745	4.371	1.267	0.591	9.385	11.320
0.25	0.25	6	1.360	2.197	3.563	2.704	14.703	16.033
1	0.25	1	5.822	8.904	0.141	0.006	6.960	9.704
1	0.25	3	3.503	5.121	1.089	0.403	9.656	11.525
1	0.25	6	1.767	2.806	3.305	2.345	14.741	16.267

Tandem systems with part-related interruptions: Numerical results for these systems are presented in Table 6. 14. For the sake of brevity, the demand arrival rate was again set to one in all the cases. The desired utilization was obtained by modifying the service times. The service distribution was deterministic and the delay distribution was exponential at all the stages. The probability of a part being delayed was set at 0.1 which is the same at all the stages. Three inter-arrival distributions in combination with three basestock levels gave a total of nine combinations. The other machine parameters are given in Table 6.13.

Table 6. 13: Machine Parameters for Systems with Part-Related Interruptions

	Stage 1	Stage 2	Stage 3
Processing time Distribution	Mean = 0.75 SCV = 0.25	Mean = 0.70 SCV = 0	Mean = 0.65 SCV = 0
Probability of delay	0.1	0.1	0.1
Downtime Distribution	Exponential Mean = 0.05	Exponential Mean = 0.10	Exponential Mean = 0.15

The results in Table 6.14 indicate that the analytical results are very close to the simulation results when the basestock level is low (S = 1) and high (S = 6). Even in the case when basestock level is 3, the analytical model captures the system behavior when the inter-arrival process is changed. Based on the limited number of cases examined, it appears that the analytical model captures the part-related interruptions better than the machine-related interruptions.

Table 6. 14: Results for Systems with Part-Related Interruptions

System Parameters		Mean Backorders at Stage 3		Average Inventory at Stage 3		Average Intermediate Inventory	
c_{0a}^2	Basestock Level	Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
0.25	1	4.164	3.969	0.070	0.806	4.225	4.108
0.25	3	1.403	0.994	1.027	1.450	6.726	7.212
0.25	6	0.284	0.142	3.644	3.978	12.117	12.964
1.00	1	6.495	6.462	0.076	0.951	4.225	4.108
1.00	3	3.289	4.474	0.792	2.115	6.726	7.212
1.00	6	1.287	1.485	2.894	3.400	12.117	12.964
2.25	1	10.178	6.854	0.085	0.001	7.350	5.682
2.25	3	6.903	3.156	0.635	0.229	11.110	7.755
2.25	6	3.639	1.605	2.303	2.143	14.693	13.291

6.9 CHAPTER SUMMARY

The decomposition framework developed in Chapters 4 and 5 was used to model the tandem system with other manufacturing features such as multiple servers, batch processing, systems with a limited supply of raw materials, multiple part type systems, and service interruptions. The numerical results indicate that further investigation is required for features like service interruptions caused by machine to develop better approximations. In modeling the other manufacturing features, the analytical approach was quite accurate in many of the cases investigated.

In the next chapter, we demonstrate how the decomposition framework can be easily extended to handle non-tandem configurations.

CHAPTER VII

MODELING FEED-FORWARD SYSTEMS AND TANDEM SYSTEMS WITH FEEDBACK

7.1 CHAPTER OVERVIEW

In this chapter, we extend the decomposition approach to two important network configurations. One is the feed-forward network which is the subject of Section 7.2, and other is a tandem system with feedback which is modeled in Section 7.3.

7.2 FEED-FORWARD SYSTEMS

In this section, a class of systems called feed-forward systems or sequential refinement systems is discussed. It is shown that the decomposition framework used in the analysis of the tandem systems can be applied to these types of systems as well.

A system of this type operates much like a specially structured network of queues. Planned inventories occur at the output of each stage which serve as an input to one or more stages. These inputs are transformed into output parts which in turn serve as input to one or more stages in front. An example of this system is shown in Figure 7.1. The input-output relationship among the stages forms a hierarchy or a tree. There is a root stage which processes raw parts and the output parts from this stage serve as input to all the other stages in the hierarchy. Each part produced at a stage may be used as input to produce any of several others, but every stage has a unique predecessor. These systems are also called refinement systems because items become gradually more specialized as they progress through processing stages. The terminal stages of the network meet the external demand and each stage joined with its predecessor, the predecessor's predecessor, etc., constitutes a path which produces a unique item. The operations are sequential which eliminates any part feedback.

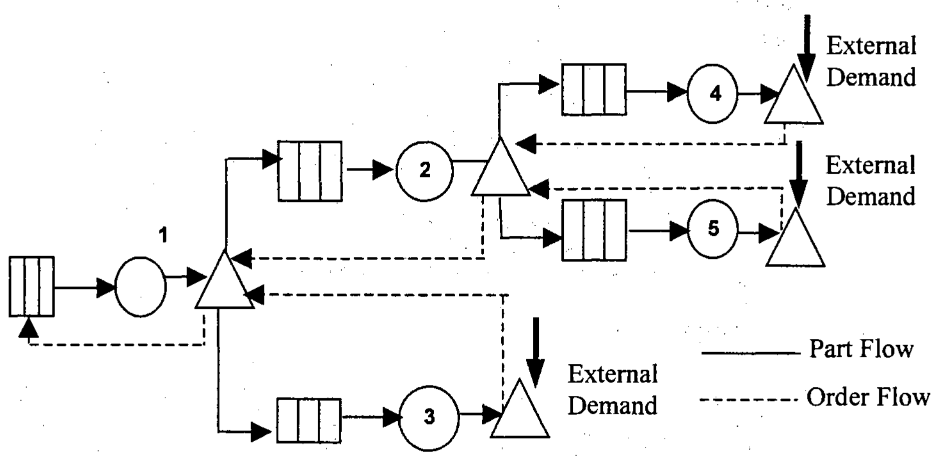


Figure 7. 1: A Feed-Forward Make-to-Stock System

Customer demands occur at the end stages. These exogenous demands in turn generate implicit demands back along the path of the predecessor items. Given these demands, each item's inventory is controlled by a local policy, specified by the basestock level. As mentioned in the description of the tandem system, this policy parameter determines the maximum planned inventory of the item.

The focus of our analysis is on systems where the arrivals follow a renewal process and processing times are general, both characterized by the first two moments of their distributions. It is also assumed that there is a single server processing parts at every stage in the system. Lee and Zipkin (1995) studied the special case of Poisson arrivals and exponential processing times.

7.2.1 Overview of the Decomposition Approach

The system consists of M stages with stage 1 being the root node. The analysis procedure begins by the computation of the demand arrival rate and SCV of the inter-arrival distribution at each stage. Let O = {M-k, M-k+1, ..., M}, k > 1, be the set of stages which meet exogenous demand. Let I = {1, 2, ..., k-1} be the set of stages whose output parts serve as input to other stages. Exogenous demands trigger the intrinsic demand at the stages in set I. The demand at a stage in set I is the superposition of the demand processes at each of its successor stages. The demand arrival rate to a stage in set I is the sum of the arrival rates at each of its successor stages. The demand inter-arrival SCV for a stage in set I is the weighted sum of inter-arrival SCVs of the successor stages. The details of the calculation are presented later in this chapter. Once demand arrival parameters have been computed, the analysis procedure is similar to the approximation procedure for a tandem system. The entire system is decomposed into one single-stage make-to-stock system (corresponding to the root node) and M-1 single-stage make-to-stock systems each with a delay node. The procedure differs in the calculation of the average number at the delay node which in turn is used in the computation of the probability distribution of the number in system in the delay model. Also, the procedure to calculate the inter-arrival time SCV to the processing node in the delay model is different from the procedure used for tandem systems as any of the node's predecessor could be a predecessor for many other nodes. The detailed analysis of a single-stage system with a delay node is not discussed here and the reader is referred to

Chapter 5. The procedure to determine the SCV of the arrival process to the processing node in the delay model is discussed in next sub-section.

7.2.2 Calculation of Demand Arrival Rates and SCVs of Demand Inter-Arrival Times

We use the following notation

M = the total number of stages;

O = the set of stages that meet external demand;

$F(i)$ = the set of stages which are the successors of stage i;

$g(i)$ = the predecessor stage for i;

λ_{0k} = exogenous demand arrival rate to stage k, $k \in$ O;

c_{0k}^2 = the SCV of the exogenous demand inter-arrival time distribution at stage k, $k \in$ O;

λ_i = demand arrival rate at stage i;

c_i^2 = the SCV of the demand inter-arrival time distribution at stage i;

c_{ai}^2 = the SCV of the inter-arrival time distribution at processing queue of stage i;

τ_i = mean processing time at stage i; and

c_{si}^2 = SCV of processing time distribution at stage i.

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

For every node k in set O,

$\lambda_k = \lambda_{0k}$; $c_k^2 = c_{0k}^2$.

For every node k in set I,

$$\lambda_k = \sum_{j \in F(k)} \lambda_j \; ; \; c_k^2 = \sum_{j \in F(k)} \frac{c_{0j}^2 . \lambda_j}{\sum_{j \in F(k)} \lambda_j} \tag{7.1}$$

This concludes the procedure for determining the demand arrival rate and SCV of demand inter-arrival distribution at a stage.

7.2.3 Determining SCV of the Inter-Arrival Process to Any Stage

The focus is now on the input side of a stage where a portion of the demand arrival process that finds a part immediately merges with a portion of the departure process from the predecessor stage. Recall that an M-stage system is decomposed into one single stage make-to-stock system (corresponding to the root node or stage 1) and M-1 single-stage make-to-stock systems each with a delay node. The delay node essentially captures the delay due to unavailability of parts at the output stores of the upstream stages. In order to analyze the single-stage system with a delay node, the SCV of the arrival process to the processing node has to be computed. The arrival rate is the same as the demand arrival rate to the node as every demand is satisfied. As demand arrives, some of the orders wait due to unavailability of parts at the predecessor stage. This affects the variability in the arrival process and, the splitting, merging and departure process approximations used by Whitt (1983) are employed to determine the effective SCV of the arrival process to the processing node. The analysis is similar to that described in Section 5.3.2 for a tandem system.

In order to determine the SCV of the arrival process at a processing node, we shift our focus to examine the departure and splitting processes at stage *i* and merging process at one of its successor stages. Figure 7.2 shows these three processes between stage *i* and one of its successors, say stage *j*. At stage *i*, the departure process splits into two, namely, the process that satisfies the backorders at all of its successor stages and the process that satisfies the replenishment orders at the output store. In other words, a portion of the orders that finish processing at stage *i*, satisfy the backorders at stage *k* for all *k*, $k \in F(i)$ and the rest of the orders proceed to the output store. Thus the portion of departure process at stage *i* that goes to satisfy backorders is further split into *n* processes, where *n* is the total number of successor stages of stage *i*.

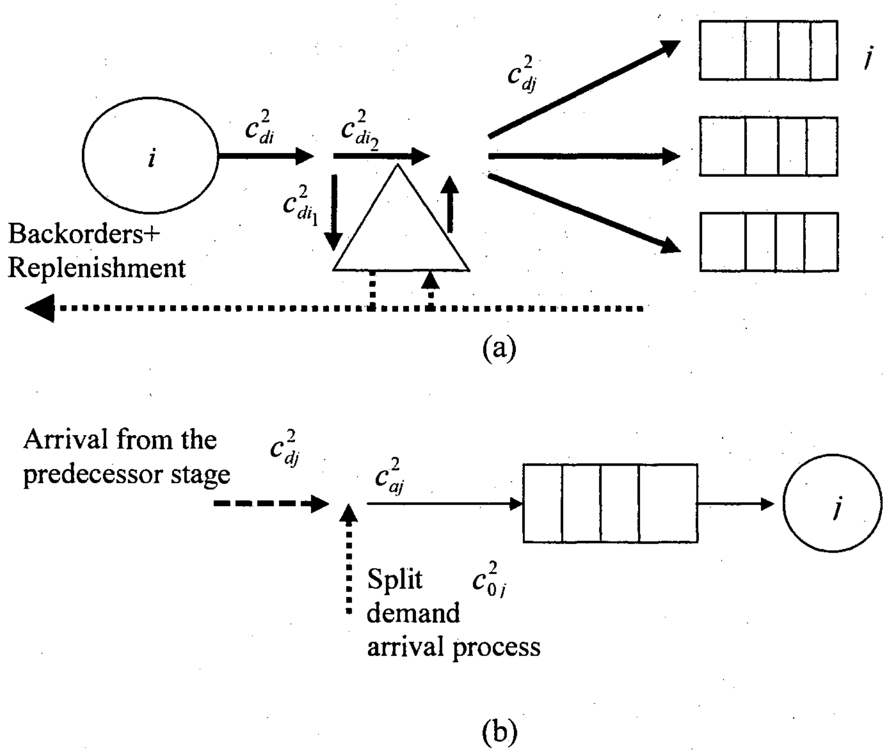


Figure 7. 2: (a) Splitting at a Predecessor Stage; (b) Merging at a Successor Stage

Looking at the arrival process at stage *j* which is one of the successors of stage i, it is the merging of the two processes, namely, those orders which obtain parts at the output store and proceed directly into processing and those which are backordered indicated by those orders which proceed from the predecessor stage *i*. We use the following procedure to determine the SCV of the arrival process at each stage. It should be noted that all of the information pertaining to stage i are known at this juncture including the arrival process to stage i.

The departure SCV at node i is given by

$$c_{di}^2 \cong \rho_i^2 c_{si}^2 + (1-\rho_i^2)c_{ai}^2 \qquad (7.2)$$

where ρ_i is the utilization of stage i.

Using the splitting approximation from Whitt (1983), SCV of the departure process that goes to the successor stages is given by

$$c_{di_2}^2 = p_i c_{di}^2 + 1 - p_i \qquad (7.3)$$

where p_i is the probability of a demand being backordered which was approximated by the probability that there is no inventory at the output store of stage i.

p_{ij}, the proportion of the departure process that goes directly to successor j is given by

$$p_{ij} = \frac{\lambda_j}{\sum_{k \in F(i)} \lambda_k}. \qquad (7.4)$$

The SCV of the arrival process from stage i to successor stage j is given by

$$c_{dj}^2 \cong p_{ij} c_{di_2}^2 + 1 - p_{ij}. \qquad (7.5)$$

Again, using the splitting approximation from Whitt (1983), we have for the split demand process

$$c_{0j}^2 = (1-p_i).c_j^2 + p_i. \qquad (7.6)$$

Using the method of superposition from Whitt (1983), the SCV of the arrival process at stage j is

$$c_{aj}^2 \cong p_i c_{dj}^2 + (1 - p_i).c_{0j}^2 \,. \tag{7.7}$$

The SCV thus determined is used as the SCV of the arrival process at the processing node in the analysis of a single-stage make-to-stock system with a delay node.

7.2.4 Analysis of Feed-Forward Make-to-Stock Systems

The solution procedure begins at the root node where it is assumed that raw materials are always available. It then proceeds with the analysis of successor nodes of the root node. It is then continued with each of the successor's successor nodes and so on until all the stages are completed. The performance measures at the root node are computed using the results from Buzacott and Shanthikumar (1993) for a single-stage make-to-stock system. Beginning at stage 2, each of the remaining M-1 stages is modeled as a single-stage make-to-stock system with a delay node. At each stage other than the first stage, we need to know the probability that an order proceeds to the delay node to compute the inter-arrival SCV to the processing node. This probability is approximated by the probability that an order will not find a part in the output store of the predecessor stage. Also, the average number at the delay node is a portion of the average of the number of backorders at the predecessor stage. At any stage backorders are created by any or all of the successor stages. Thus, the proportion of backorders corresponding to a successor stage is equal to the ratio of demand arrival rate from that node to the sum of the demand arrival rates from all of the successor stages. Once, the average number at the delay node is obtained, all the performance measures for that node can be computed using the delay model.

The complete procedure is presented in an algorithmic form. In the following, $i = 1, 2, \ldots, M$.

Stage 1 (Root Node):

The expected number in system and the expected number of backorders in this stage are given by,

$$\mathrm{E}[\mathrm{N}_1] = \lambda.\mathrm{E}[\mathrm{W}_{\mathrm{q}1}] + \rho_1;\; E[B_1] \cong \frac{\rho_1 \sigma_1^{S_1}}{(1-\sigma_1)}, \text{ where } \sigma_1 = \frac{E[N_1]-\rho_1}{E[N_1]}. \quad (7.8)$$

$\mathrm{E}[W_{q1}]$ is the expected waiting time in queue at stage 1 and is calculated using the GI/G/1 approximation presented in Chapter 5. The expression for $\mathrm{E}[\mathrm{B}_1]$ is obtained from the analysis of a GI/G/1 make-to-stock system contained in Buzacott and Shanthikumar (1993).

Also, we approximate the probability that a demand from a successor stage will not find a part in the output store using

$$p_1 \cong \rho_1 . \sigma_1^{S-1}. \quad (7.9)$$

Stage *i* ($i \in$ F(1))

The SCV of the arrival process $\mathrm{c}_{\mathrm{ai}}^2$ is calculated using the procedure given in Section 7.2.3. The average number at stage *i* is the sum of the orders corresponding to stage *i* waiting for parts from the predecessor stage *g(i)*, and the orders with parts waiting to complete processing (including the one in process). Note the *g(i)* = 1, initially.

Hence,

$$E[N_i] = E[B_{g(i)}].\frac{\lambda_i}{\sum_{k \in F(i)} \lambda_k} + \lambda_i E[W_{qi}] + \rho_i. \tag{7.10}$$

Let ρ_d be the average number at the delay node. ρ_d is given by

$$E[B_{g(i)}].\frac{\lambda_i}{\sum_{k \in F(i)} \lambda_k}. \tag{7.11}$$

Note that ρ_d is simply the portion of the average number of backorders that corresponds to orders from stage *i*. Using the single-stage model with delay node, the steady state probability that there are *n* parts in the output store is given by,

$$P[I_i = n] \cong \begin{cases} e^{-\rho_d} \frac{\rho_d^{S_i - n}}{(S_i - n)!}(1-\rho_i) + \sum_{j=1}^{S_i - n} e^{-\rho_d} \frac{\rho_d^{S_i - n - j}}{(S_i - n - j)!}\rho_i .(1-\sigma_i).\sigma_i^{j-1} & n = 1, 2, \ldots, S_i - \\ e^{-\rho_d}.(1-\rho_i) & n = S_i \end{cases} \tag{7.12}$$

and the expected inventory at stage *i* is simply

$$\mathrm{E[I_i]} = \sum_{\mathrm{n=1}}^{\mathrm{S_i}} \mathrm{n.P[I_i = n]}. \tag{7.13}$$

The expected backorders at stage *i* is given by,

$$\mathrm{E[B_i] = E[N_i] + E[I_i] - S_i} \tag{7.14}$$

The probability that a demand from any of the successor stages will not find a part in the output store of stage *i* is approximated by the steady state probability that the output store of stage i is empty. That is,

$$p_i \cong 1 - \sum_{n=1}^{S_i} P[I_i = n]. \tag{7.15}$$

After the procedure is completed for all the successor stages of the root stage 1, then the above procedure is continued with the successor stages of each of successor stages of stage 1 and so on.

7.2.5 Numerical Results

A three-stage system with stage 1 being the root stage, stage 2 and stage 3 being the terminal stages, satisfying external demand, was used as the test system. The parameters for the various configurations are given in Table 7.1. The external mean demand arrival rate was set to one as in the previous cases. Two demand inter-arrival distributions were used, namely, Erlang and exponential. Tables 7.2 and 7.3 compare the average number of backorders (stages 2 and 3), the average inventory level (stages 2 and 3), and the average intermediate inventory for the simulation and analytical models. The results indicate that the approximation is reasonably accurate in all the configurations tested.

Table 7. 1: Feed-Forward Network Configurations

Configuration	Stage 1 Service Distribution SCV	Stage 1 Basestock Level	Stage 2 Service Distribution SCV	Stage 2 Basestock Level	Stage 3 Service Distribution SCV	Stage 3 Basestock Level
1	0.25	3	1.00	6	2.25	9
2	0.25	3	2.25	9	1.00	6
3	1.00	6	0.25	3	2.25	9
4	1.00	6	2.25	9	0.25	3
5	2.25	9	0.25	3	1.00	6
6	2.25	9	1.00	6	0.25	3

Table 7. 2: Feed-Forward Systems: Exponential Inter-Arrival Times at Stage 2 and Exponential Inter-Arrival Times at Stage 3

	Average Backorders at Stage 3		Average Backorders at Stage 2	
Configuration	Simulation	Analytical	Simulation	Analytical
1	1.739	1.789	1.187	1.156
2	1.202	1.245	1.854	1.755
3	1.928	1.776	1.349	1.258
4	1.308	1.291	1.951	1.794
5	1.586	1.269	1.840	1.456
6	1.773	1.419	1.676	1.322

	Average Inventory at Stage 3		Average Inventory at Stage 2		Average Intermediate Inventory	
Configuration	Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
1	4.461	4.280	2.836	2.697	10.843	11.168
2	2.871	2.658	4.458	4.296	10.949	11.246
3	4.338	4.278	1.061	0.934	11.802	11.822
4	1.075	0.928	4.340	4.270	11.873	11.887
5	2.662	2.470	1.000	0.786	11.766	11.469
6	0.983	0.791	2.617	2.450	11.809	11.500

Table 7. 3: Feed-forward Systems: Erlang Inter-Arrival Times at Stage 2 and Exponential Inter-Arrival Times at Stage 3

	Average Backorders at Stage 3		Average Backorders at Stage 2	
Configuration	Simulation	Analytical	Simulation	Analytical
1	0.900	1.106	1.037	1.101
2	0.337	0.549	1.891	1.705
3	0.779	1.063	1.190	1.161
4	0.298	0.383	1.681	1.738
5	0.622	0.569	1.543	1.332
6	0.707	0.504	1.378	1.245

	Average Inventory at Stage 3		Average Inventory at Stage 2		Average Intermediate Inventory	
Configuration	Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
1	5.013	4.819	3.000	2.861	9.747	10.364
2	3.464	3.214	4.505	4.465	10.060	10.412
3	5.111	4.812	1.110	1.035	10.389	11.013
4	1.429	1.282	4.580	1.413	10.649	11.063
5	3.161	2.973	1.041	0.880	10.671	10.685
6	1.292	1.081	2.760	2.592	10.726	10.713

7.3 TANDEM SYSTEMS WITH FEEDBACK

In all of the systems studied so far in this dissertation, order-processing flow was restricted to be in one direction. Also, the parts after completing processing were always assumed to be of perfect quality. In this section, a class of systems is considered where some limited feedback is allowed that could be used to model the possibility of producing imperfect quality parts. The details of the feedback process are explained in the next section. Zipkin (1995b) studied a special case of these systems where the demand arrival process is Poisson and the processing times are exponential

7.3.1 System Description

Demand for finished goods arrive as a renewal process to stage M, characterized by its mean and SCV, and is always for a single part. Each stage has an inventory of processed parts at the output side which in turn is used to make parts at the next stage. The dynamics of the system is similar to the tandem make-to-stock system described in Chapters 4 and 5. However, the processing at each stage can produce defective parts. If a defective part is produced, it is discarded and an order is sent to the preceding stage to obtain another finished part from the preceding stage to compensate for the discarded part. It is assumed that feedback is allowed only to the immediate predecessor stage. The following parameters are defined for the model. In the following, j = 1, 2, …, M.

λ = External demand arrival rate;

c_0^2 = SCV of the demand inter-arrival distribution;

τ_j = mean processing time at stage *j*;

c_{sj}^2 = SCV of the processing time distribution at stage *j*; and

r_j = the probability that a unit at stage j passes to the next stage.

The key observation in this system is that at each stage, if a defective part is produced, the part is scrapped and additional demand is generated since the request for replenishment/backorder is not satisfied yet. *The same part/customer is not re-circulated.* $(1- r_j)$ indicates the portion of parts that are scrapped, and the orders are then routed back to its predecessor stage to obtain fresh parts. The demand arrival to any node i is a superposition of external demand and internally generated demand resulting from feedback. Since our decomposition approach proceeds sequentially from stage 1, we will not be able to compute the variability parameters of the combined demand arrival process, because we do yet have information about the SCV of the internal demand process generated by the downstream stages. To approximately handle this situation, two approaches were developed.

Method 1: We first computed the total demand rate at each stage and then assumed that the SCV of the combined demand arrival process was the same as that of the external demand arrival process. Then, the sequential decomposition approach was used to analyze the system.

Method 2: First, method 1 was executed to solve the complete system. This gave us values (although incorrect) for the SCVs of the departure process from all the stages. We used these to update the SCVs of the combined demand arrival process at each stage. At the completion of processing at a stage, the departure process first splits into two. One part that corresponds to good parts proceeds to satisfy demand. The other that corresponds to discarded parts now becomes the internal demand that is fed back to the previous stage. We now solve the tandem line with the updated demand arrival process parameters. This gives a new set of departure process SCVs and the above procedure is repeated until there is no appreciable change in the SCV values.

7.3.2 Computation of Total Demand Arrival Rates

Let λ_i denote the total demand rate at stage *i*. It is given by

$$\lambda_i = \frac{\lambda}{r_M . r_{M-1}, \ldots, r_i}. \tag{7.16}$$

Now using this total demand arrival rate, the approximation method for tandem make-to-stock systems is used to calculate the system performance measures.

7.3.3 Numerical Results

Three-stage systems were used for testing the approximation. Let ***R*** denote the feedback probability vector where r_i is the probability that a part at stage *i* proceeds to the next stage. Two different ***R*** vectors were used in combination with parameters from the non-homogeneous systems used earlier (see Table 5.10). A total of 12 different configurations were tested. The arrival process is Poisson with a rate of one.

The two different R vectors were

$$R_1 = \begin{bmatrix} 1 \\ 0.8 \\ 0.9 \end{bmatrix}; \text{ and } R_2 = \begin{bmatrix} 1 \\ 1 \\ 0.9 \end{bmatrix}$$

The results obtained using the approximation described in Method 1 is presented in Tables 7.4 and 7.5. The tables show that the results from the analytical model are very close to the simulation models for most of the configurations tested. Ten out of the twelve cases examined had less than 12% RPD for average intermediate inventory.

Table 7. 4: Tandem Make-to-Stock Systems with Feedback: Feedback Vector is R_1

	Average Backorders at Stage 3		Average Inventory at Stage 3		Average Intermediate Inventory	
Configuration	Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
1	1.173	0.846	5.050	4.446	10.146	10.517
2	2.268	0.926	2.928	1.595	13.384	13.793
3	1.450	0.462	4.612	4.160	9.291	9.145
4	3.596	2.315	1.093	0.210	14.794	14.947
5	2.897	0.882	2.284	1.374	10.778	9.650
6	3.849	2.057	1.036	0.248	13.060	11.951

Table 7. 5: Tandem Make-to-Stock Systems with Feedback: Feedback Vector is R_2

	Average Backorders at Stage 3		Average Inventory at Stage 3		Average Intermediate Inventory	
Configuration	Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
1	0.273	0.269	6.339	6.239	10.354	10.530
2	0.304	0.202	3.996	3.928	12.737	12.774
3	0.267	0.273	5.727	6.156	11.428	10.117
4	0.531	0.320	1.619	1.550	14.843	14.769
5	0.304	0.200	3.953	3.898	13.916	11.469
6	0.483	0.302	1.638	1.602	9.896	13.867

To see if the iterative method, Method 2, improved the accuracy, analysis was performed using the R_2 vector for all the six configurations. The sequential procedure was rerun ten times. After each execution, the arrival process to stage 3 was modified by the superposition of the split departure process from stage 3 and the external arrival process. Since, there was no feedback to stages 2 and 1, the modified inter-arrival SCV calculated for stage 3 was also used for stages 2 and 1. The results are presented in Table 7.6.

The results indicate that there was a very marginal improvement in the average inventory at stage 3 and the average intermediate inventory in some cases. The average number of backorders did not show any improvement. A more extensive numerical study is needed before any further conclusions can be drawn.

Table 7. 6: Results using the Approximation in Method 2

Configuration	Average Backorders at Stage 3		Average Inventory at Stage 3		Average Intermediate Inventory	
	Simulation	Analytical	Simulation	Analytical	Simulation	Analytical
1	0.273	0.257	6.339	6.269	10.354	10.488
2	0.304	0.195	3.996	3.943	12.737	12.753
3	0.267	0.260	5.727	6.200	11.428	10.060
4	0.531	0.308	1.619	1.560	14.843	14.748
5	0.304	0.186	3.953	3.941	13.916	11.412
6	0.483	0.279	1.638	1.623	9.896	13.822

7.4 CHAPTER SUMMARY

In this chapter, we showed that the decomposition framework that was applied to tandem configurations in earlier chapters can be extended to certain non-tandem configurations. The two configurations modeled, feed-forward networks and tandem lines with feed-back are two of the most commonly found configurations in the real world systems. The analytical results were reasonably accurate in many of the cases examined.

The next and final chapter summarizes the contributions of this dissertation and identifies directions for future work.

CHAPTER VIII

SUMMARY, CONCLUSIONS AND FUTURE RESEARCH

First, we summarize the research carried out in this dissertation effort. This is followed by a summary of the research contributions that were made to the performance analysis body of knowledge. We conclude this chapter by identifying some directions for future research.

8.1 RESEARCH SUMMARY

The main research goal of this dissertation was to develop analytical models for production-inventory systems where queueing, inventory and reliability issues can be simultaneously addressed. In Chapter 4, a sequential decomposition approach for analyzing tandem make-to-stock production systems with Poisson arrivals and exponential processing times was developed where the queueing and inventory issues were addressed within the same framework. The numerical results indicated that the analytical approximation performed very accurately, and better than published methods. The approach was generalized to address general arrival processes and general service time distributions in Chapter 5. Several configurations were tested and results indicated that the approximation method performed extremely well in most cases. Chapter 6 extended the decomposition approach to model additional manufacturing features such as multiple servers, batch processing of parts, limited supply of raw materials, multiple part type systems, and systems with service interruptions. This led to a framework being established wherein the decomposition procedure was used to model these additional manufacturing features. The results showed that approximation performed well in many of the configurations tested. This demonstrated that the framework was versatile in handling additional manufacturing features including some reliability and quality characteristics. In Chapter 7, the same framework was then generalized to model feed-forward type networks and tandem make-to-stock systems with limited feedback. Some additional contributions are presented in the appendices. Efforts to develop improved approximations for the basic tandem system are summarized in Appendix A.2. Appendix A.4 explores the applicability of performability analysis to production-inventory systems. It also explores an alternative approach that uses a stochastic Petri net model that includes performance and reliability issues within a single unified model.

8.2 RESEARCH CONTRIBUTIONS

The primary contribution of this research was the development of an analytical modeling framework that can simultaneously address inventory and capacity/congestion issues in a wide variety of production-inventory systems. With regard to modeling of reliability issues, it was shown that the framework could handle issues such as service interruptions and product quality.

The various contributions are summarized below.

- The modeling power of the parametric decomposition approach based on the two-moment queueing framework was extended by including inventory issues.

- A general performance analysis framework was established based on the newly developed sequential decomposition procedure for analyzing tandem make-to-stock systems. The framework was used to model several manufacturing features like multiple servers, batch processing, limited supply of raw materials and multiple-part types.
- By modeling feed-forward networks and some limited types of feedback, it was shown that the approach developed has the potential to handle general system configurations.
- The applicability of the performability analysis framework in production-inventory systems was shown.
- The models developed could be the foundation for developing more comprehensive performance analysis models of supply chain networks.

By building on existing analytical models and methods, this research has shown that a more unified and comprehensive analytical approach can be developed for the performance analysis of manufacturing systems.

8.3 FUTURE DIRECTIONS

It is always true that a solution to a research problem leads to many other interesting problems which remain to be solved. This dissertation is no exception, and some of key future directions are pointed out.

An assumption in all of the systems examined was that every demand that arrived was satisfied. Systems with limited backorders or no backorders resemble Kanban systems where the number of Kanban cards limits the number of orders in the system. Extension of the decomposition framework in modeling Kanban systems could be explored.

Another assumption was the one-for-one replenishment inventory policy that was used at all of the stages. Modeling of other inventory policies could be a subject of future research.

Example systems where many of the manufacturing features are simultaneously present could be investigated to test the robustness of the decomposition approach.

While modeling service interruptions, the models did not yield accurate results in comparison with the simulation results. Further investigation is required to develop better approximations for this feature.

Feed-back mechanism was modeled only in the context of tandem systems. Modeling feedback in other configurations would be the next step in extending the framework to a general network configuration.

REFERENCES

Altiok, T. (1989), "(R,r) Production/Inventory Systems," *Operations Research*, 38, 2, 266-276.

Altiok, T. and R. Ranjan (1995), "Multi-Stage, Pull-type Production/Inventory Systems," *IIE Transactions*, 26, 4, 190-200.

Altiok, T. and G. A. Shiue (1994), "Single-Stage, Multi-product Production/Inventory Systems with Backorders," *IIE Transactions*, 26, 1, 266-276.

Askin, R. G. and C. R. Standridge (1993), *Modeling and Analysis of Manufacturing Systems*, John Wiley & Sons, Inc., New York, NY.

Bitran, G.R. and S. Dasu (1992), "A Review of Open Queueing Network Models of Manufacturing Systems," *Queueing Systems: Theory and Applications*, 12, 95-134.

Bitran, G. R. and D. Tirupati (1989), "Approximations for Product Departures from a Single Server Station with Batch Processing in Multi-Product Queues," *Management Science*, 35, 7, 851-878.

Buzacott, J. A. (1989), "Queueing Models of Kanban and MRP Controlled Production Systems," *Engineering Costs and Production Economics*, 17, 3-20.

Buzacott, J. A. and J. G. Shanthikumar (1992), "A General Approach for Coordinating Production in Multiple-Cell Manufacturing Systems," *Production and Operations Management*, 1, 1, 34-52.

Buzacott, J. A. and J. G. Shanthikumar (1993), *Stochastic Models of Manufacturing Systems*, Prentice Hall, Inc., Englewood Cliffs, NJ.

Ciardo, G., J. Muppala and K. Trivedi (1989), "SPNP: Stochastic Petri Net Package," *Third International Work-Shop on Petri Nets and Performance Models*, Kyoto, Japan, 142-151.

Donatiello, L. and B. R. Iyer (1987), "Analysis of a Composite Performance Reliability Measure for Fault-Tolerant Systems," *Journal of the Association of Computing Machinery*, 34, 179-199.

Federgruen, A. (1993), "Centralized Planning Models for Multi-Echelon Inventory Systems under Uncertainty," In *Handbooks in Operations Research and Management Science*, Volume 4, Logistics of Production and Inventory, S. C. Graves, A. H. G. Rinnooy Kan and P. H. Zipkin, Eds., North-Holland, Amsterdam, The Netherlands, 133-171.

Gavish, B. and S. Graves (1980), "A One-Product Production/Inventory Problem under Continuous Review Policy," *Operations Research*, 28, 2, 1228-1236.

Gershwin, S. B. (1994), *Manufacturing Systems Engineering*, Prentice Hall, Inc., Englewood Cliffs, NJ.

Goyal, S. K. and A. Gunasekaran (1990), "Multi-Stage Production-Inventory Systems," *European Journal of Operational Research*, 46, 1, 1-20.

Iyer, B.R., L. Donatiello and P. Hiedelberger (1986), "Analysis of Performability for Stochastic Models of Fault-Tolerant Systems," *IEEE Transactions on Computers*, 35, 902-907.

Jackson, J. R. (1957), "Network of Waiting Lines," *Operations Research*, 5, 518-521.

Jackson, J. R. (1963), "Job-Shop like Queueing Systems," *Management Science*, 10, 131-142.

Jagerman, D. L. (1984), "Methods in Traffic Calculations," *AT&T Bell Laboratories Technical Journal*, 63, 1283-1310.

Kamath, M. (1994), "Recent Developments in Modeling and Performance Analysis Tools for Manufacturing Systems," In *Computer Control of Flexible Manufacturing Systems: Research and Development*, S.B. Joshi and J. S. Smith, Eds., Chapman & Hall, London, UK, 231-263.

Koenigsberg, E. (1959), "Production Lines and Internal Storage- A Review," *Management Science*, 5, 401-433.

Kraemer W. and M. Langenbach-Belz (1976), "Approximate Formulae for the Delay in the Queueing System GI/G/1," *Proceedings of the Eighth International Teletraffic Congress*, Melbourne, 235-(1-8).

Kuehn, P. J. (1979), "Approximate Analysis of General Queueing Networks by Decomposition," *IEEE Transactions on Communication*, 27, 113-126.

Law, A. M. and W. D. Kelton (1991), *Simulation and Modeling & Analysis*, Second Edition, McGraw-Hill, Inc., New York, NY.

Lee, H. L. and C. Billington (1993), "Material Management in Decentralized Supply Chains," *Operations Research*, 41, 5, 835-847.

Lee, H. L. and C. Billington (1995), "The Evolution of Supply-Chain-Management Models and Practice at Hewlett-Packard," *Interfaces*, 25, 5, 42-63.

Lee, Y. J. and P.H. Zipkin (1992), "Tandem Queues with Planned Inventories," *Operations Research*, 40, 5, 936-947.

Lee, Y. J. and P.H. Zipkin (1995), "Processing Networks with Inventories: Sequential Refinement Systems," *Operations Research*, 43, 6, 1025-1036.

Little, J. D. C. (1961), " A Proof of the Queueing Formula: $L = \lambda.W$," *Operations Research*, 9, 383-387.

Meyer, J. F. (1980) "On Evaluating the Performability of Degradable Computing Systems," *IEEE Transactions on Computers*, 29, 8, 720-731.

Morse, P. M. (1958), *Queues, Inventories and Maintenance: The Analysis of Operational Systems with Variable Demand and Supply*, John Wiley and Sons, New York, NY.

Muckstadt, J. A. and R. O. Roundy (1993), "Analysis of Multistage Production Systems," In *Handbooks in Operations Research and Management Science*, Volume 4, Logistics of Production and Inventory, S. C. Graves, A. H. G. Rinnooy Kan and P. H. Zipkin, Eds., North-Holland, Amsterdam, The Netherlands, 59-129.

Pattipati, K. R. (1993), "A Unified Framework for the Performability Evaluation of Fault-Tolerant Computer Systems," *IEEE Transactions on Computers*, 42, 3, 312-325.

Pritsker, A. A. B. (1995), *Introduction to Simulation and SLAM II*, Fourth Edition, John Wiley and Sons, Inc., New York.

Ram, R. and N. Viswanadham (1995), "Performability of Automated Manufacturing Systems with Centralized Material Handling," *International Journal of Production Research*, 32, 8, 1775-1799.

Segal, M. and W. Whitt (1989), "A Queueing Network Analyzer for Manufacturing," *Proceedings of the 12th International Teletraffic Congress*, 1146-1152.

Shanthikumar, J. G. and J. A. Buzacott (1980), "On the Approximations to the Single Server Queue," *International Journal of Production Research*, 18, 761-763.

Sherbrooke, C. C. (1968), "METRIC: A Multi-Echelon Technique for Recoverable Item Control," *Operations Research*, 16, 122-141.

Sivaramakrishnan S. and M. Kamath (1996), "Performance Evaluation of Production-Inventory Systems," at *INFORMS Atlanta Fall 1996 Meeting*, November 3 - November 6, 1996.

Sivaramakrishnan, S. and M. Kamath (1997), "Analytical Models of Multi-Stage Make-to-Stock Systems," *6th Industrial Engineering Research Conference Proceedings*, IIE, Norcross, GA, 795-800.

Smith, R. A, K. S. Trivedi and A. V. Ramesh (1988), "Performability Analysis: Measures, an Algorithm and a Case Study," *IEEE Transactions on Computers*, 37, 4, 407-417

Suri, R. (1988), "A New Perspective on Manufacturing Systems Analysis," In *Design and Analysis of Integrated Manufacturing Systems*, W. D. Compton, Ed. National Academy Press, Washington, DC, 118-133.

Suri, R. and S. de Treville (1991), "Full Speed Ahead: A Timely Look at Rapid Modeling Technology in Operations Management," *OR/MS Today*, June, 34-42.

Suri, R. and S. de Treville (1993), "Rapid Modeling: The Use of Queueing Models to Support Time-Based Competitive Manufacturing," In *Operations Research in Production Planning and Control*, Springer-Verlag, 21-30.

Suri, R. and G. W. Diehl (1987), "Rough-cut Modeling An Alternative to Simulation," *CIM Review*, 3, 25-32.

Suri, R., G. W. Diehl and R. Dean (1986), "Quick and Easy Manufacturing Systems Analysis Using MANUPLAN," *Proceedings of the Spring IIE Conference*, IIE, 195-205

Suri, R., G. W. Diehl, S. de Treville and M. J. Tomsicek (1995), "From CAN-Q to MPX: Evolution of Queueing Software for Manufacturing," *Interfaces*, 25, 5, 128-150.

Suri, R, J. L. Sanders and M. Kamath (1993), "Performance Evaluation of Production Networks," In *Handbooks in Operations Research and Management Science*, Volume 4, Logistics of Production and Inventory, S. C. Graves, A. H. G. Rinnooy Kan and P. H. Zipkin, Eds. North-Holland, Amsterdam, The Netherlands, 59-129.

Svoronos, A. and P. H. Zipkin (1991), "Evaluation of One-for-One Replenishment Policies for Multi-Echelon inventory systems," *Management Science*, Vol. 37, No. 1, pp. 68-83.

Tirupati, D. (1992), "Application of the Parametric Decomposition Approach in the Analysis of Manufacturing Queueing Networks: An Overview," *of the 1st Industrial Engineering Research Conference Proceedings*, IIE, Norcross, GA, 307-311.

Viswanadham, N. and Y. Narahari (1992), *Performance Modeling of Automated Manufacturing Systems*, Prentice-Hall, Inc., Englewood Cliffs, NJ.

Viswanadham, N., Y. Narahari and R. Ram (1991), "Performability of Automated Manufacturing Systems," *Control and Dynamic systems*, 47, 77-119.

Viswanadham, N., K. Pattipati and V. Krishna (1995), "Performability studies of Automated Manufacturing Systems with Multiple-Part Types," *IEEE Transactions on Robotics and Automation*, 11, 5, 692-709.

Welch, P. D. (1983), "The Statistical Analysis of Simulation Results," In *Computer Performance Modeling Handbook*, S. S. Lavenberg, Ed., Academic Press.

Whitt, W. (1983), "The Queueing Network Analyzer," *Bell Systems Technical Journal* 62, 779-815.

Whitt, W. (1993), "Approximations for the GI/G/m Queue," *Production and Operations Management*, 2, 2, 114-154.

Wolff, R. W. (1982), "Poisson Arrivals See Time Averages," *Operations Research*, 30, 223-231.

Zheng, Y. and P. H. Zipkin (1990), "A Queueing Model to Analyze the Value of Centralized Inventory Information," *Operations Research*, 38, 2, 296-307.

Zipkin, P. H. (1984), "Models for Design and Control of Stochastic, Multi-Item Batch Production Systems," *Operations Research*, 34, 1, 91-104.

Zipkin, P. H. (1995a), "Performance Analysis of a Multi-Item Production-Inventory System Under Alternative Policies," *Management Science*, 41, 4, 690-703.

Zipkin, P. H. (1995b), "Processing Networks with Planned Inventories: Tandem Queues with Feedback," *European Journal of Operational Research*, 80, 344-349.

APPENDIX A1

DETERMINATION OF WARM-UP PERIOD AND RUN LENGTH FOR THE SIMULATION EXPERIMENTS

A1.1 INTRODUCTION

In this dissertation, the accuracy of the analytical results was determined by comparison with simulation estimates. The performance measures obtained from analytical models represent steady-state values of system behavior. Thus, the simulation estimates obtained must represent steady-state performance measures. While performing the steady-state simulation experiments, a warm-up period has to be determined to remove any initialization bias, and sufficient run length should be provided so that infrequent events occur a reasonable number of times.

In this study, a wide variety of systems were tested with simulation estimates. It would have been extremely time consuming to determine a warm-up period and run length for each and every system configuration. On the other hand, it is necessary to determine a proper warm-up period and run length to obtain statistically accurate simulation estimates. The factors that affect warm-up period and run length are the stochastic parameters that describe the system such as the processing time and demand inter-arrival time parameters. In general, the higher the variability in the stochastic components, the longer it would take for the system to reach steady state and longer would be the run length to get good estimates. A similar statement can be made with respect to the utilization level. Thus, the system with the highest utilization and the highest variance in the stochastic components was chosen to determine warm-up and run-length for all the simulation experiments.

A three stage tandem system was chosen. Demand inter-arrival time and the processing times at each stage followed a hyper-exponential distribution (SCV = 2.25). The mean of the inter-arrival time was one and the mean processing times was 0.80, thus a utilization of 0.80 was obtained at each stage. The base stock level at each stage was set at zero. Each simulation run was terminated after 16,000 time units.

A1.3 WARM-UP PERIOD AND RUN LENGTH DETERMINATION

The general technique developed by Welch (1983) was used to determine the warm-up period for the system described above. The procedure is described briefly, next.

n replications of the simulation (n = 20 for this system) are made, each of length m (m > 14,000). Let Y_{ij} represent the time in system for the i^{th} observation from the j^{th} replication, (i = 1, 2, ..., m) and (j = 1, 2, ..., n).

Let $\overline{Y}_i = \sum_{j=1}^{n} \frac{Y_{ij}}{n}$ for i = 1, 2, ..., m. The averaged process $\overline{Y}_1$, $\overline{Y}_2$, ... has means $E(\overline{Y}_i) = E(Y_i)$ and variances $\mathrm{Var}(\overline{Y}_i) = \mathrm{Var}(Y_i)/n$. Thus, the averaged process has the same transient mean curve as the original process, but its plot has only $(1/n)^{th}$ of the variance.

To smooth out the high frequency oscillations in $\overline{Y}_1$, $\overline{Y}_2$, ..., the moving average $\overline{Y}_i(w)$, where w is the window and is a positive integer is calculated as follows:

$$\bar{Y}_i(w) = \begin{cases} \sum_{s=-w}^{w} \frac{\bar{Y}_{i+s}}{2w+1} & \text{if } i = w+1, \ldots, m-w \\ \sum_{s=-(i-1)}^{i-1} \frac{\bar{Y}_{i+s}}{2i-1} & \text{if } i = 1, 2, \ldots, w \end{cases}$$

Thus, if i is not too close to the beginning of the replications, then $\bar{Y}_i(w)$ is just the simple average of $2w + 1$ observations of the averaged process centered at observation i. It is called a moving average since i moves through time.

$\bar{Y}_i(w)$ is plotted for $i = 1, 2, \ldots, m\text{-}w$ and the value of i beyond which $\bar{Y}_1(w)$, $\bar{Y}_2(w)$, ... appears to converge is identified and becomes the warm-up period.

The above procedure was applied to the system described in Section A.1.1. The values of n and m were based on the recommendations of Law and Kelton (1991). n was chosen to be 20 and m was set at 16,000 parts. The plot of $\bar{Y}_i(w)$ is shown in Figures A.1.1 through A.1.4. The measure plotted is the average time in system.

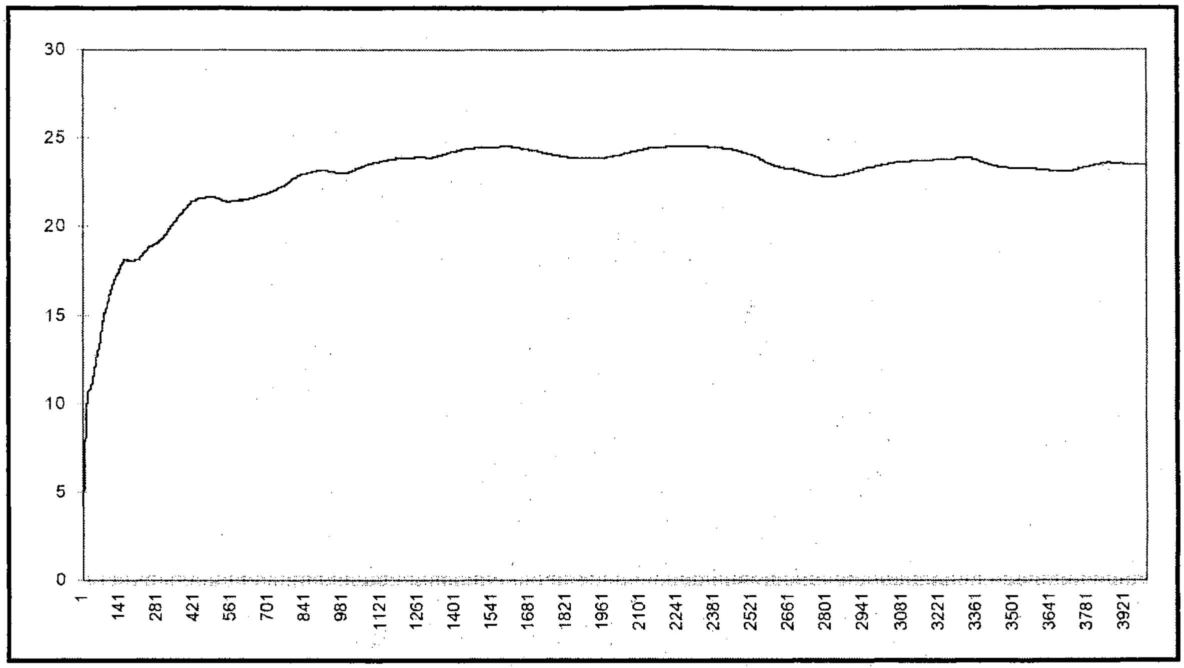


Figure A1. 1: The plot of $\overline{Y}_i(w)$

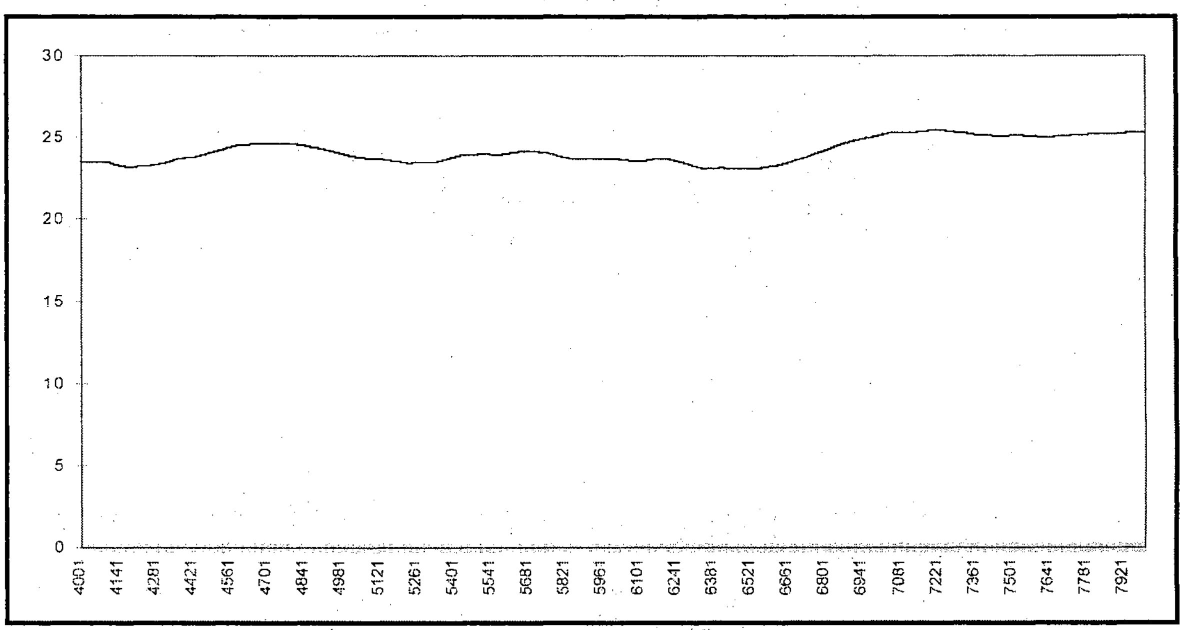


Figure A1. 2: The plot of $\overline{Y}_i(w)$ (Continued)

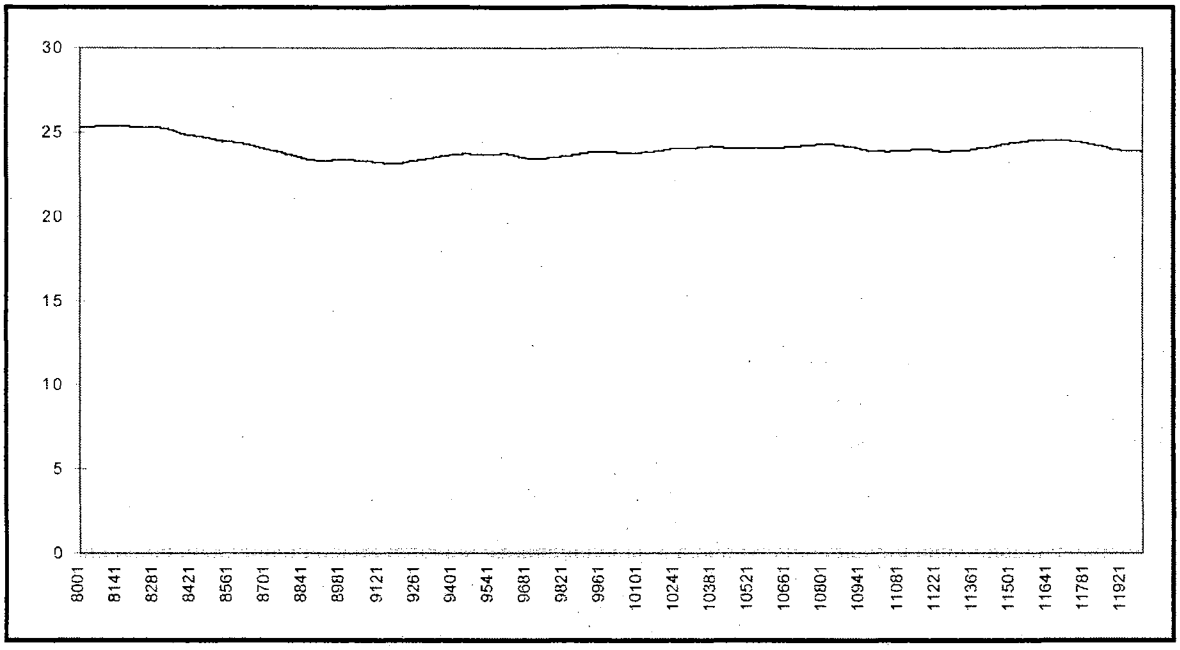


Figure A1. 3: The plot of $\overline{Y}_i(w)$ (Continued)

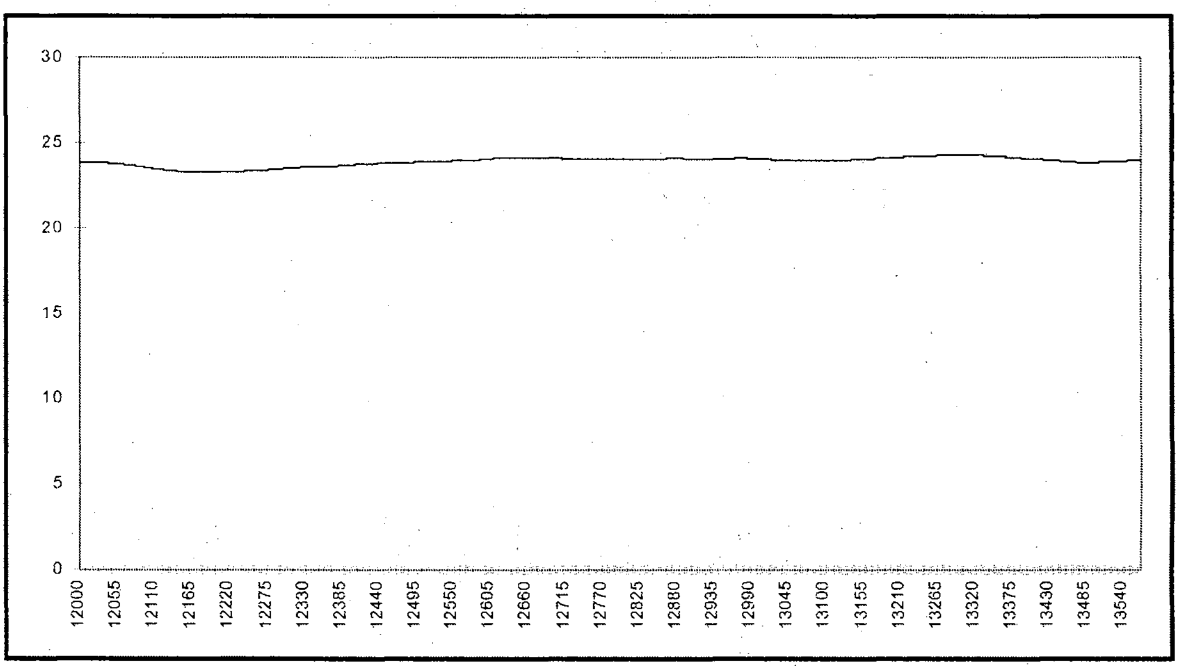


Figure A1. 4: The plot of $\overline{Y}_i(w)$ (Continued)

From Figures A.1.1 through A.1.4, the measure seems to stabilize after processing of about 2,000 products with the moving average of time in system varying between 23 time units and 25 time units. After the completion of 12,000 products, the variation in the moving average reduces and stabilizes at about 24 time units. Three different warm-up periods of 5,000, 10,000 and 15,000 with run lengths after warm-up of 50,000, 100,000 and 150,000 time units were tested to determine a suitable combination of warm-up period and run length.

Another decision that was required in combination with warm-up period and run length was the random number seeds that should be chosen such that the sampled data from random number generators are independent of each other as well as between runs. In SLAM II, 10 different random number streams are available. Any simulation program coded in SLAM II is limited to combinations of these 10 streams. Also, the user can choose the starting unnormalized random number seed for any of the 10 random number streams, and if the seeds are not specified, the SLAM II processor uses default values. Several combinations of user specified or default seeds together with single or multiple random number streams were used to decide on the warm-up period and run length for the experiments.

Table A1. 1: Mean and 95% confidence interval for a 3-Stage System with Hyper-Exponential Inter-Arrival and Service Distributions; $\rho = 0.80$

SEEDS SET BY	***STREAMS USED***	***WARM-UP = 5,000***	***WARM-UP = 10,000***	***WARM-UP = 15,000***
SLAM	ONE (Run length after warm-up 50,000)	24.88 ± 0.805	24.77 ± 0.924	23.94 ± 0.548
SLAM	ONE (Run length after warm-up 100,000)	24.85 ± 0.356	24.44 ± 0.582	25.12 ± 0.709
SLAM	ONE (Run length after warm-up 150,000)	24.75 ± 0.312	24.75 ± 0.391	24.80 ± 0.380
USER	ONE (Run length after warm-up 50,000)	25.00 ± 0.824	24.96 ± 0.822	25.09 ± 0.944
USER	ONE (Run length after warm-up 100,000)	25.05 ± 0.431	24.99 ± 0.504	25.04 ± 0.552
USER	SEVEN (Run length after warm-up 50,000)	24.10 ± 0.559	24.19 ± 0.616	24.32 ± 0.728
SLAM	SEVEN (Run length after warm-up 50,000)	25.04 ± 0.717	24.84 ± 0.661	24.38 ± 0.632

Each experiment was replicated ten times and the mean time in system for a customer along with the 95% confidence interval was computed. Table A1.1 gives the various combinations along with the interval estimates. The average time in system for the various combinations varies from 23.94 to 25.12. The configuration with a warm-up of 5,000, with a single seed set by SLAM and a run-length of 50,000 after warm-up was selected for the all the simulation experiments, as it was not significantly different from the other configurations. The number of replications was set at 10, which were large enough to get a half-width which was less than 5% of the mean value.

APPENDIX A2

IMPROVING THE ACCURACY OF THE DECOMPOSITION APPROACH

The extensive numerical investigation carried out in Chapter 5, indicated that there is a need to improve the accuracy of the approximation for some of the cases. In this appendix, we describe all the attempts that were made to improve the accuracy of the approximation procedure. We approached this problem by examining the key assumptions made in our approximation scheme.

A2.1 IMPROVING THE DELAY MODEL

A single-stage make-to-stock with a delay which is called the delay model is a key building block in our approach. The delay node is modeled using the $M/G/\infty$ queue in our approach. When the external demand arrival process is not Poisson, we still use an $M/G/\infty$ model instead of a $GI/G/\infty$ model as exact expressions are not available for the latter case. Hence, as a first step, we investigate if an approximate solution to the $GI/G/\infty$ system produced better overall results than the exact solution of the $M/G/\infty$ (approximate) model. Whitt (1993, 94) suggested the use of a normal approximation to obtain the probability of the number of busy servers in $GI/G/\infty$ system. We examined the accuracy of this approximation with simulation results and also compared it with the results from the $M/G/\infty$ model.

Table A2.1 shows the comparison of the probability distribution of number in system in a GI/G/∞ queue with hyper-exponential inter-arrival time and Erlang processing times using simulation, normal approximation and the M/G/∞ approximation. The normal approximation seems to approximate the distribution better than the M/G/∞ approximation. The next step was to implement this approximation in various make-to-stock systems, and compare the results with both simulation and the original approximation. However, within the larger scope of the approximation for the analysis of the make-to-stock system, the GI/G/∞ approximation did not significantly improve the overall results. Hence, it was decided that the M/G/∞ approximation would be used for the delay model in all cases.

Table A2. 1: Probability Distribution of the Number in System in a GI/G/∞ Queue with Hyper-Exponential Inter-Arrival Times and Erlang Service Times
Mean Number in System = 4.0

Number in System	Simulation	Normal Approximation	M/G/∞ Approximation
0	0.069	0.1075	0.0183
1	0.121	0.0520	0.0733
2	0.149	0.0803	0.1465
3	0.151	0.1095	0.1954
4	0.138	0.1319	0.1954
5	0.114	0.1410	0.1563
6	0.089	0.1319	0.1042
7	0.063	0.1095	0.0595
8	0.043	0.0803	0.0298
9	0.028	0.0520	0.0132
10	0.017	0.0298	0.0053
11	0.009	0.0150	0.0019
12	0.005	0.0067	0.0006
13	0.003	0.0026	0.0002
14	0.001	0.0010	0.0000
15	0.001	0.0003	0.0000

A2.2 A MODIFIED DECOMPOSITION APPROACH

The next attempt to improve the approximation explored an alternative way of determining the distribution of the number of backorders. The basic idea was to substitute the single-stage delay network with an *equivalent* GI/G/1 make-to-stock system. The expected number of backorders obtained from this equivalent GI/G/1 system was used as an estimate of the expected number of backorders in the single-stage delay model. The detailed procedure is described next for a tandem system with single-server stages.

Stage 1 is a GI/G/1 make-to-stock system because of our assumption that raw materials are always available. Using the approximations contained in Buzacott and Shanthikumar (1993) for a GI/G/1 make-to-stock system, all of the steady state measures can be obtained for stage 1 as shown later in this section.

At stage 2, the total number of orders is the sum of orders that are waiting for parts from the previous stage (stage 1) and orders with parts that are waiting or being processed at the this stage. In other words, the total number at any stage is the sum of backorders at the previous stage and orders that are waiting or being processed. In the original approximation presented in Chapters 4 and 5, we modeled this stage using a single-stage delay model. In this new method, we replace the delay model with an equivalent GI/G/1 make-to-stock system. The expected number in this equivalent system is the same as the expected number at stage 2. Using Buzacott and Shanthikumar's (1993) approximation for a GI/G/1 make-to-stock system, the expected number of backorders are calculated. The rest of the performance measures are easily obtained using standard relationships. This procedure is repeated for all the M-1 stages. The procedure described in Chapter 5, Section 5.3.2 is used to compute the rate and SCV of arrivals to the processing stage. The complete procedure is presented next in an algorithm form. The notation defined in Chapter 5 is followed.

Stage 1:

The expected number in system and the expected number of backorders in this stage are given by,

$$E[N_1] = \lambda . E[W_{q1}] + \rho_1 ; \; E[B_1] = \frac{\rho_1 \sigma_1^{S_1}}{(1-\sigma_1)}, \text{ where } \sigma_1 = \frac{E[N_1] - \rho_1}{E[N_1]} \quad \text{(A2.1)}$$

$E[W_{q1}]$ is the expected waiting time in queue at stage 1 and it is calculated using Kraemer and Langenbach-Belz (1976) approximation, which is given by

$$E[W_{q1}] \cong g\left(\frac{c_{a1}^2 + c_{s1}^2}{2}\right)\left(\frac{\tau_1 \rho_1}{(1-\rho_1)}\right) \quad \text{(A2.2)}$$

where $g \equiv g(\rho_1, c_{a1}^2, c_{s1}^2)$ is defined as

$$g(\rho_1, c_{a1}^2, c_{s1}^2) = \begin{cases} \exp\left(\dfrac{-2(1-\rho_1)(1-c_{a1}^2)^2}{3\rho_p (c_{a1}^2 + c_{s1}^2)}\right), & c_{a1}^2 < 1 \\ \exp\left(\dfrac{-(1-\rho_1)(c_{a1}^2 - 1)}{(c_{a1}^2 + 4c_{s1}^2)}\right), & c_{a1}^2 > 1 \end{cases} \quad \text{(A2.3)}$$

$$c_{a1}^2 = c_{0a}^2 \quad \text{(A2.4)}$$

The expression for $E[B_1]$ is obtained from the analysis of a GI/G/1 make-to-stock system contained in Buzacott and Shanthikumar (1993).

$$E[B_1] = \frac{\rho_1 \sigma_1^{S_1}}{1-\sigma_1} \text{ where } \sigma_1 = \frac{E[N_1] - \rho_1}{E[N_1]} \quad \text{(A2.5)}$$

$E[I_1]$ is obtained using the relationship $E[I_1] = E[B_1] + S - E[N_1]$

Stage i (i > 1):

The SCV of the arrival process c_{ai}^2 is calculated using the procedure given in Section 5.3.2. The average number at node *i* is the sum of the orders waiting for parts from the previous stage, stage *i*-1, and the orders with parts waiting to complete processing (including the one in process).

Hence, $E[N_i] = E[B_{i-1}] + \lambda . E[W_{qi}] + \rho_i$ (A2.6)

E[N$_i$] is now viewed as the expected number in system in a GI/G/1 make-to-stock system representing stage *i* and the backorders at stage *i-1*.

The expected number of backorders at this stage is given by,

$$E[B_i] = \frac{\rho_i \sigma_i^{S_i}}{1-\sigma_i} \text{ where } \sigma_i = \frac{E[N_i] - \rho_i}{E[N_i]} \quad \text{(A2.7)}$$

The expected inventory at the output store is simply

$$E[I_i] = E[B_i] + S - E[N_i]. \quad \text{(A2.8)}$$

The probability of an order at a stage not finding a part at its predecessor's stage output store is obtained from the inventory distribution in a GI/G/1 make-to-stock system (Buzacott and Shanthikumar, 1993). Beginning at stage 2, this procedure is repeated sequentially till the last stage.

A2.3 NUMERICAL RESULTS

The modified procedure was tested for many cases where the original approximation did not yield accurate results. Figures A2.1 through A2.7 compare the expected number of backorders calculated by the modified procedure and simulation estimates. Homogeneous systems were used to test the modified approximation.

It can be seen that the modified method performs better when both the arrival and service variability are high. Recall that our original approximation did not perform very well for these cases. The new method complements the original approximation in that it can be used when both the service and arrival processes have high variability.

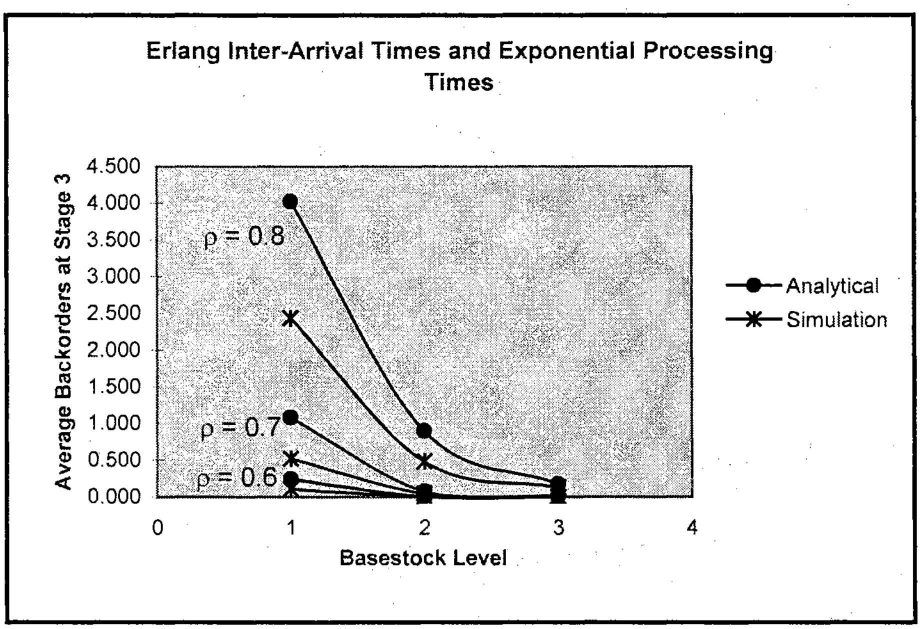


Figure A2. 1: E[B3] for Erlang Inter-Arrival Times and Exponential Processing Times

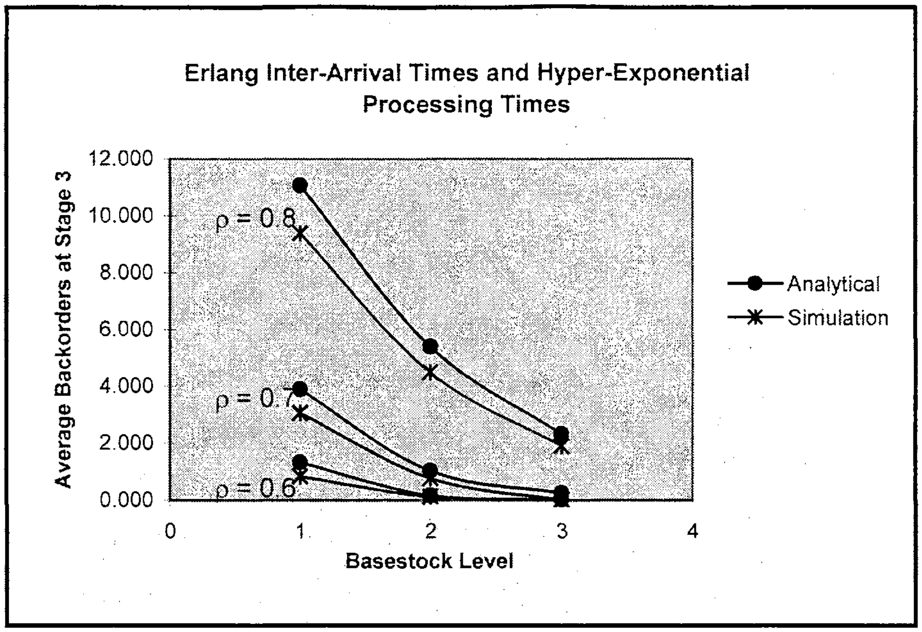


Figure A2. 2: E[B_3] for Erlang Inter-Arrival Times and Hyper-Exponential Processing Times

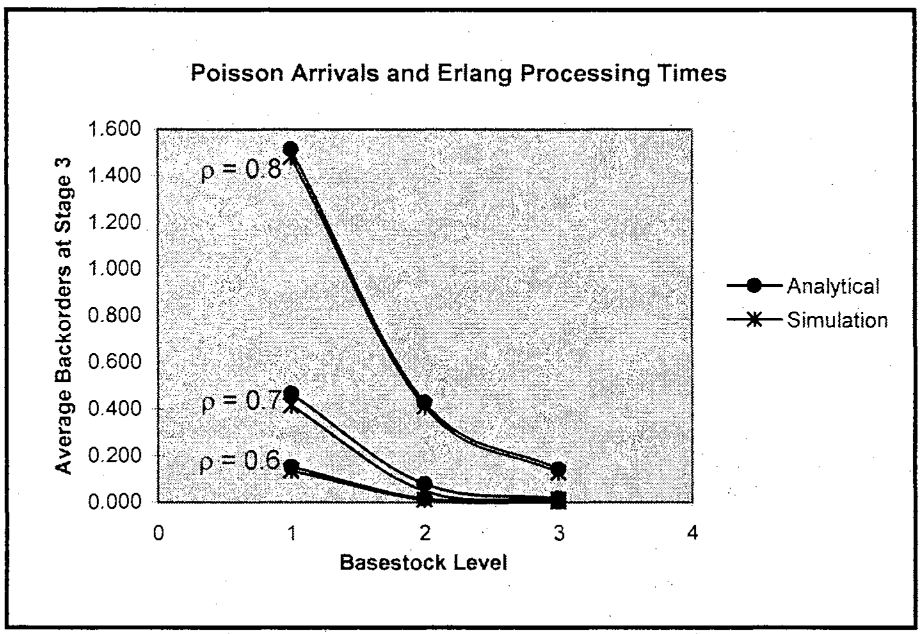


Figure A2. 3: E[B_3] for Poisson Arrivals and Erlang Processing times

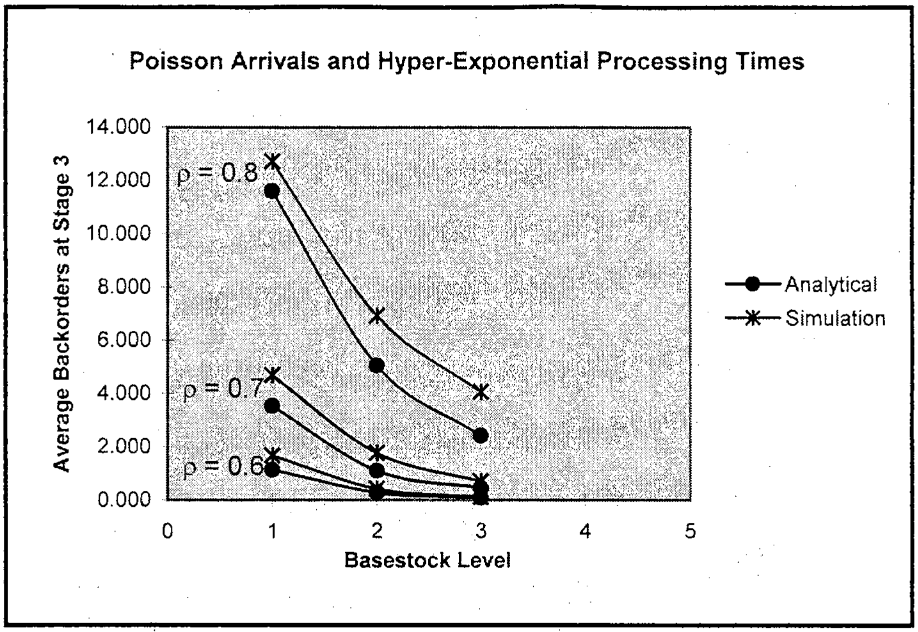


Figure A2. 4: E[B$_3$] for Poisson Arrivals and Hyper-Exponential Processing Times

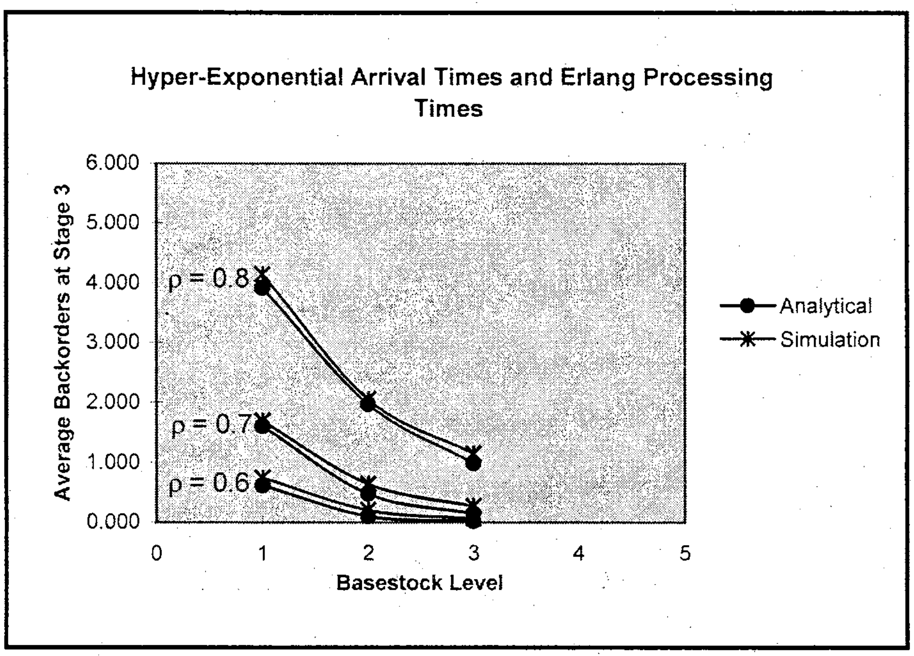


Figure A2. 5: E[B$_3$] for Hyper-Exponential Inter-Arrival Times and Erlang Processing Times

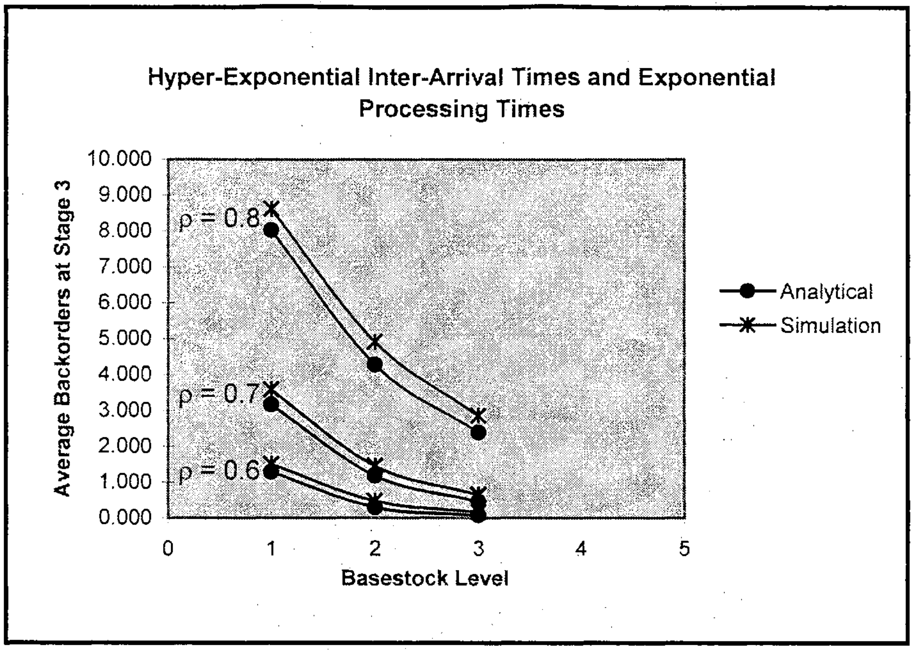


Figure A2. 6: E[B3] for Hyper-Exponential Inter-Arrival Times and Exponential Processing Times

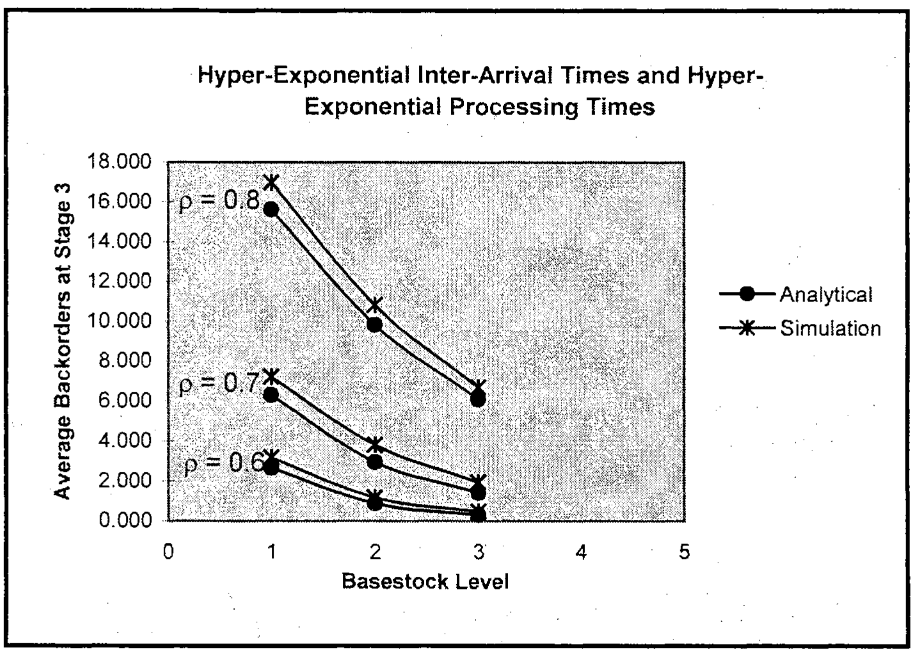


Figure A2. 7: E[B3] for Hyper-Exponential Inter-Arrival Times and Hyper-Exponential Processing Times

APPENDIX A3

DESCRIPTION OF A SAMPLE SIMULATION MODEL

In this dissertation, simulation estimates were used to test the accuracy of the analytical approximations developed. In Appendix A1, the method used for determining warm-up period and run length was described. In this section, we present the logic behind one of the simulation models used. The description presented here is for a three-stage make-to-stock system with single-server stages. This basic model served as the starting point for the other simulation models as in the case of the analytical model.

A3.1 THE SIMULATION LOGIC

The flow chart of the events is presented in Section A3.3. A total of six global variables were used, two for each stage, corresponding to the inventory and backorder levels at a stage. The XX() variables in SLAM II were used for this purpose. The program model begins with the generation of demand for finished products. XX(2) represents the current inventory level at stage 3. When XX(2) is positive, it implies that there is inventory, and thus the inventory is reduced by one and an order is triggered to replenish this satisfied demand. If XX(2) is zero, then it implies that there is no inventory and thus the demand is backordered. The variable XX(1) which represents the backorders at stage 3 is now increased by one. Any demand irrespective of whether it was satisfied immediately or not triggers an order to be processed at stage 3. The order thus generated looks into the output store of stage 2. If parts are available at the output store which is verified by XX(4) being greater than zero, the inventory at stage 2 is reduced by one and the order proceeds to join the queue for processing at stage 3. If parts are not available at stage 2, the variable XX(3) which represents the backorder level at stage 2 is increased by one. As in the case of stage 3, every demand at stage 2 triggers an order for processing. This order now follows a process similar to that in case of stage 3. The inventory level is reduced at stage 1 (XX(6)) or the backorders are increased (XX(5)) at stage 1 as the result of this process. The orders at stage 1 directly enter processing as it is assumed that raw materials are always available.

The SIMULATE and MONTR statements in SLAM II are used to initiate a run and clear statistics, respectively. The listing of the program is provided next.

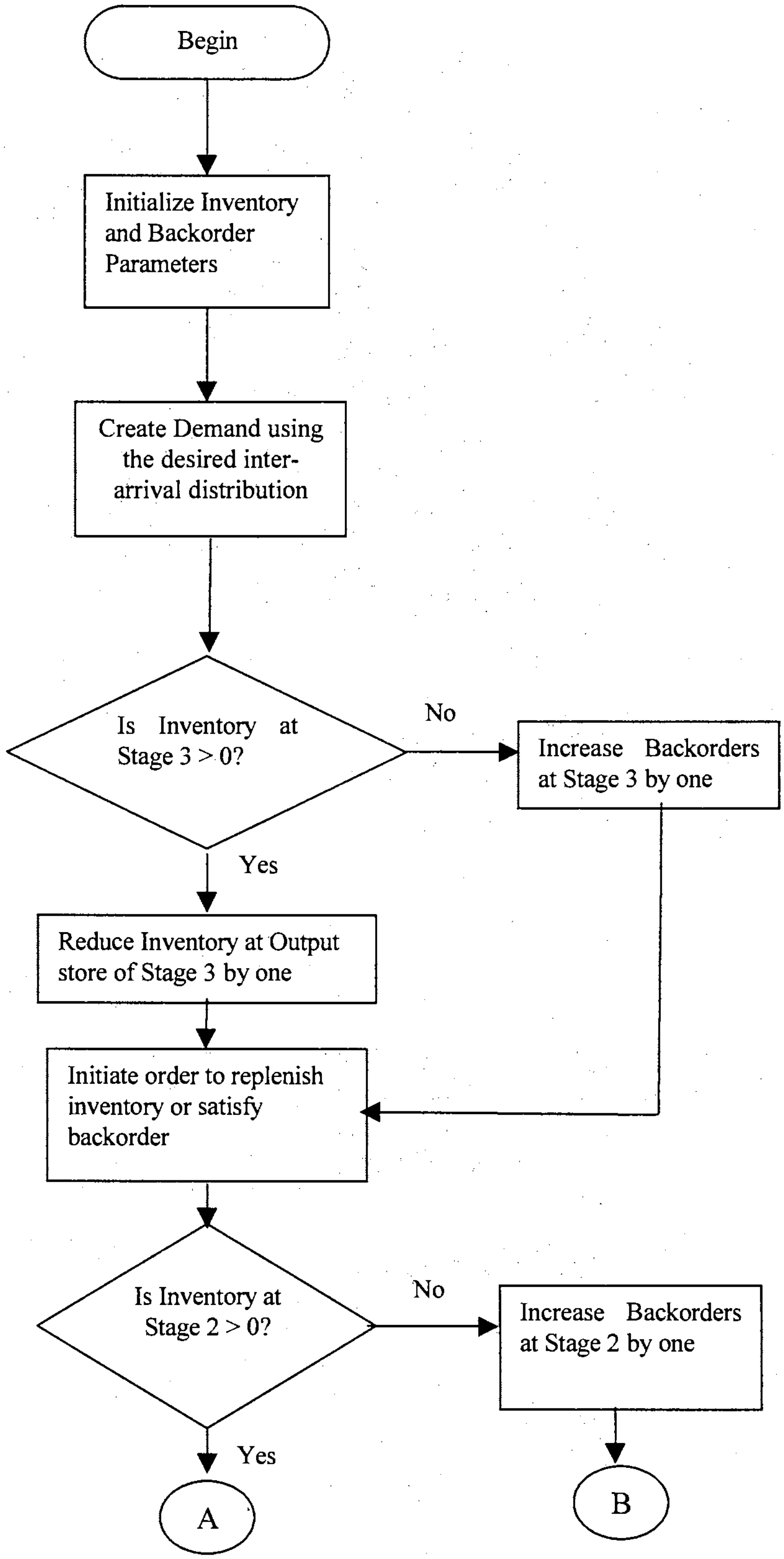
A3.2 PROGRAM LISTING

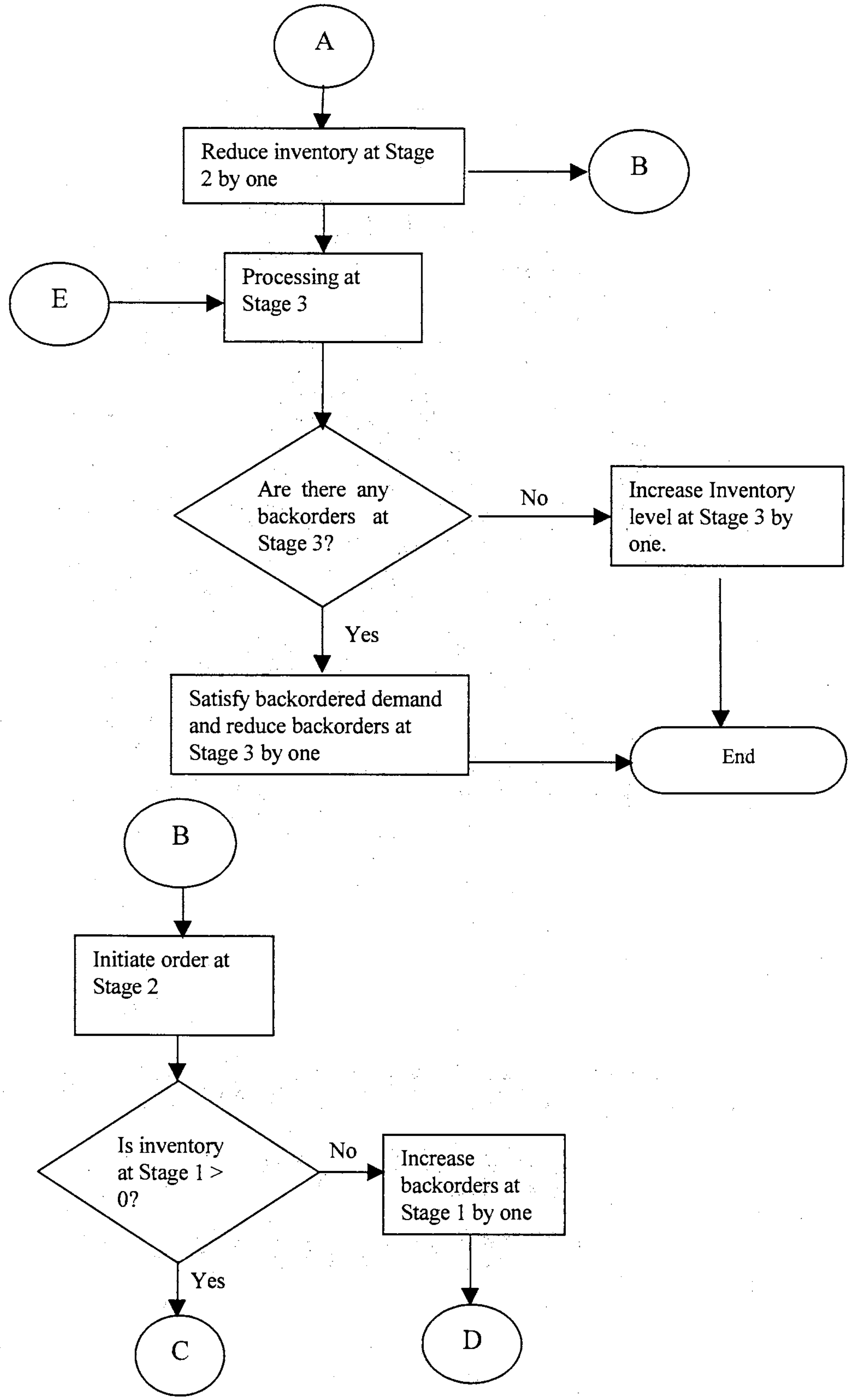
```
GEN,SHANKAR,THESIS,3/24/1997,1,N,N,Y/Y,N,Y/1,132;
LIMITS,3,2,500;
INTLC,XX(1)=0,XX(3)=0,XX(5)=0,XX(6)=3,XX(4)=3,XX(2)=3;
TIMST,XX(1),BO AT N3;
TIMST,XX(3),BO AT N2;
TIMST,XX(5),BO AT N1;
TIMST,XX(2),INVEN AT N3;
TIMST,XX(4),INVEN AT N2;
TIMST,XX(6),INVEN AT N1;
NETWORK;
      CREATE,ERLNG(0.25,4,1),1,1;
      ACTIVITY;
COL1  COLCT,BET,TOTAL ORDERS;
      ACTIVITY,,XX(2).GT.0;
      ACTIVITY,,XX(2).EQ.0,ASG2;
COL2  COLCT,BET,NUM FULFILLED;
      ACTIVITY;
ASG1  ASSIGN,XX(2)=XX(2)-1;
      ACTIVITY;
GON1  GOON,1;
      ACTIVITY,,XX(4).GT.0;
      ACTIVITY,,XX(4).EQ.0,ABO2;
WST2  ASSIGN,XX(4)=XX(4)-1;
      ACTIVITY;
```

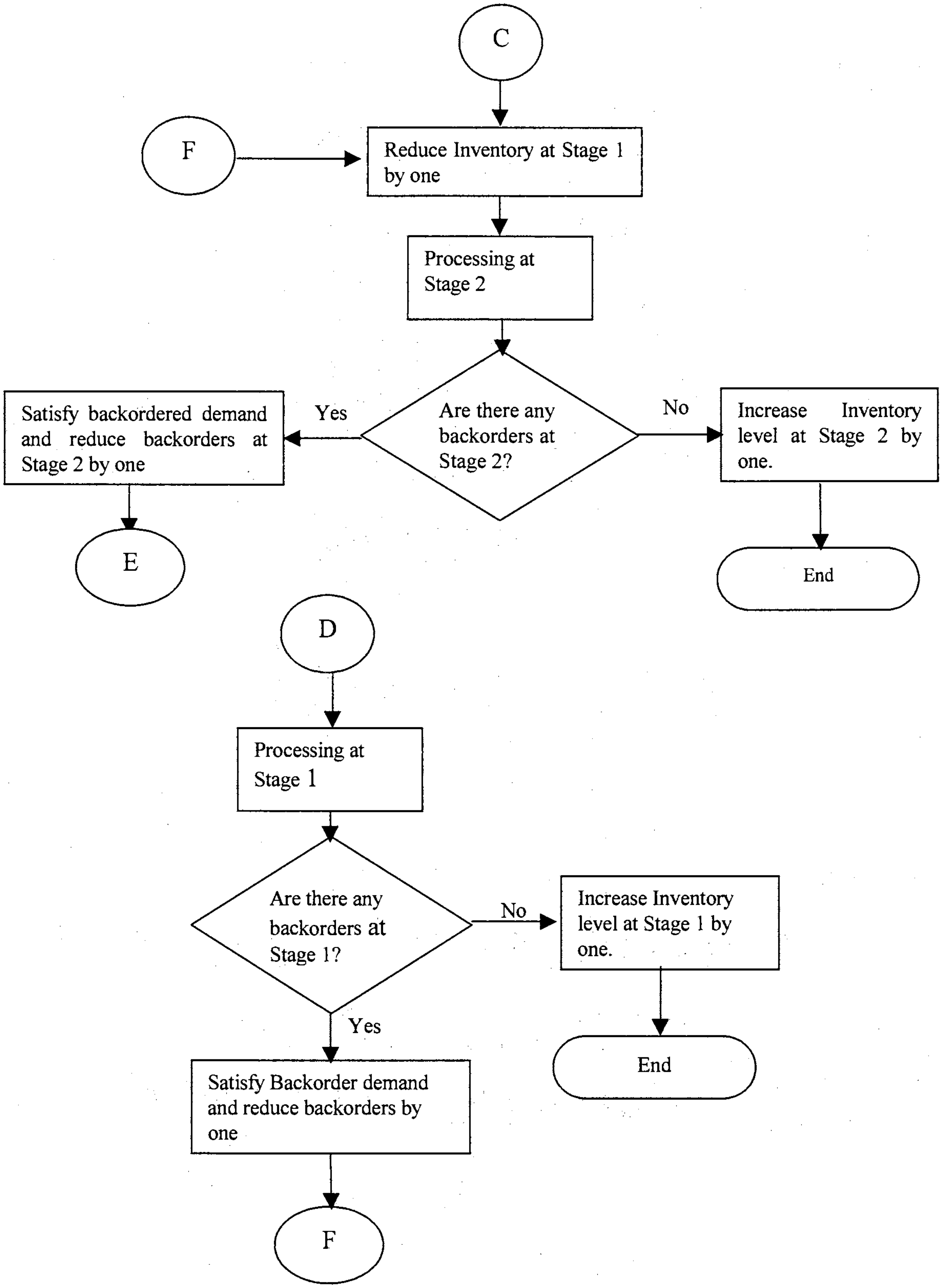
```
        ACTIVITY,,,GON2;
NOD3    COLCT,BET,ARR AT 3;
        QUEUE(1),,,;
        ACTIVITY(1),EXPON(0.8,1);
COL3    COLCT,INT(1),TIME BET ORDERS,,1;
        ACTIVITY,,XX(1).GT.0;
        ACTIVITY,,XX(1).EQ.0,STO1;
ASG5    ASSIGN,XX(1)=XX(1)-1;
        ACTIVITY;
        TERMINATE;
STO1    ASSIGN,XX(2)=XX(2)+1;
        ACTIVITY;
        TERMINATE;
GON2    GOON,1;
        ACTIVITY,,XX(6).GT.0;
        ACTIVITY,,XX(6).EQ.0,ABO3;
WST3    ASSIGN,XX(6)=XX(6)-1;
        ACTIVITY;
        ACTIVITY,,,NOD1;
NOD2    COLCT,BET,ARR AT 2;
        QUEUE(2),,,;
        ACTIVITY(1),EXPON(0.8,1);
        GOON,1;
        ACTIVITY,,XX(3).GT.0;
        ACTIVITY,,XX(3).EQ.0,STO2;
        ASSIGN,XX(3)=XX(3)-1;
        ACTIVITY,,,NOD3;
STO2    ASSIGN,XX(4)=XX(4)+1;
        ACTIVITY;
        TERMINATE;
NOD1    COLCT,BET,ARR AT 1;
        QUEUE(3),,,;
        ACTIVITY(1),EXPON(0.8,1);
        GOON,1;
        ACTIVITY,,XX(5).GT.0;
        ACTIVITY,,XX(5).EQ.0,STO3;
RBO1    ASSIGN,XX(5)=XX(5)-1;
        ACTIVITY,,,NOD2;
STO3    ASSIGN,XX(6)=XX(6)+1;
        ACTIVITY;
        TERMINATE;
ABO3    ASSIGN,XX(5)=XX(5)+1,1;
        ACTIVITY,,,NOD1;
ABO2    ASSIGN,XX(3)=XX(3)+1;
        ACTIVITY,,,GON2;
ASG2    ASSIGN,XX(1)=XX(1)+1,1;
        ACTIVITY,,,GON1;
        END;
INITIALIZE,,55000,Y;
MONTR,CLEAR,5000;
SIMULATE;
MONTR,CLEAR,5000;
SIMULATE;
MONTR,CLEAR,5000;
```

```
SIMULATE;
MONTR,CLEAR,5000;
SIMULATE;
MONTR,CLEAR,5000;
SIMULATE;
MONTR,CLEAR,5000;
SIMULATE;
MONTR,CLEAR,5000;
SIMULATE;
MONTR,CLEAR,5000;
SIMULATE;
MONTR,CLEAR,5000;
SIMULATE;
MONTR,CLEAR,5000;
SIMULATE;
FIN;
```

A3.3 FLOWCHART FOR THE SIMULATION MODEL




A
Reduce inventory at Stage 2 by one
B
E
Processing at Stage 3
Are there any backorders at Stage 3?
No
Increase Inventory level at Stage 3 by one.
Yes
Satisfy backordered demand and reduce backorders at Stage 3 by one
End
B
Initiate order at Stage 2
Is inventory at Stage 1 > 0?
No
Increase backorders at Stage 1 by one
Yes
C
D


C
F
Reduce Inventory at Stage 1 by one
Processing at Stage 2
Are there any backorders at Stage 2?
Yes
Satisfy backordered demand and reduce backorders at Stage 2 by one
E
No
Increase Inventory level at Stage 2 by one.
End
D
Processing at Stage 1
Are there any backorders at Stage 1?
No
Increase Inventory level at Stage 1 by one.
End
Yes
Satisfy Backorder demand and reduce backorders by one
F

APPENDIX A4

PERFORMABILITY ANALYSIS OF MAKE-TO-STOCK SYSTEMS

Performability analysis, a combined analysis of system reliability and system performance was described in Chapter 2. The purpose of this appendix is to show that such analysis can also be performed in the context of make-to-stock systems. We present two methods of performability analysis, one is a technique where the structure state process was used in conjunction with the performance model to derive the performability measures. This method is described in Section A4.1. The other technique involves the solution of a combined model using stochastic Petri net theory. This type of analysis is discussed in Section A4.2.

A4.1 PERFORMABILITY ANALYSIS USING THE STRUCTURE STATE PROCESS

Consider the feed-forward network shown in Figure A4.1. The dynamics of the performance model is similar to such systems discussed in Chapter 7. At each stage, the machine is prone to failures. When any of the machines at stage 2 through stage 5 fails, the system continues to work and produce parts. If the processor at stage 1 fails, the system shuts down. It is assumed that a machine once failed is not repaired within the observation period.

The performance model wherein we obtain measures like the expected number of backorders, the expected inventory level and the proportion of demand that is met immediately are obtained using the approximation developed in the earlier chapters. The reliability model is described by the structure state process having the state space {0, 1, 2, 3, 4} where the interpretation of the states is as follows:

0 : stage 1 machine failed or all other machines failed

i: stage 1 machine operating and exactly *i* machines at stage 2 through stage 5 are operating. $i = 1, 2, 3, 4$.

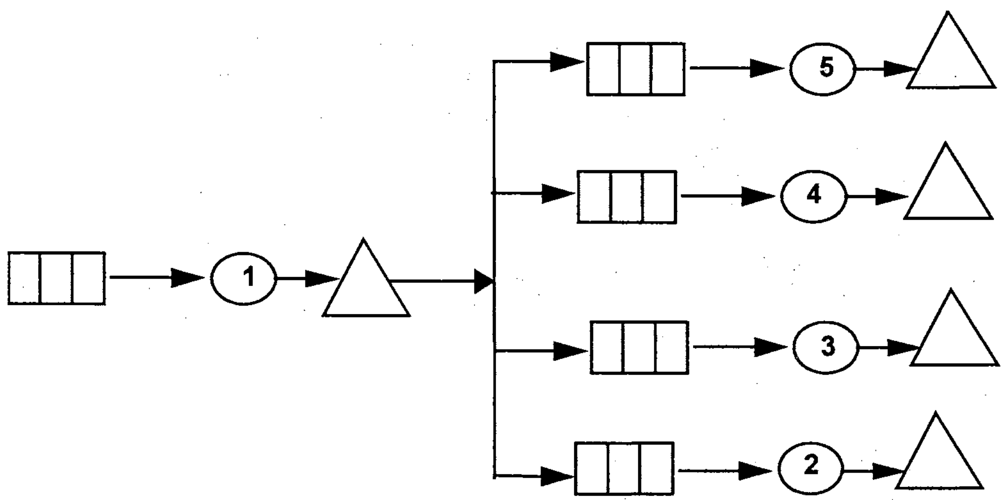


Figure A4. 1: A Feed-Forward Make-to-Stock System

Let the time to failure of the machine at stage 1 and the time to failure of the other machines be exponentially distributed with rates α_a and α, respectively. Also, let the failures be independent of one another.

The structure state process (SSP) for this system is shown in Figure A4.2, which is the same as the one considered by Donateillo and Iyer (1987). Hence, their approach could be used to find the performability distribution for the feed-forward system. Next, we present an outline of the overall approach.

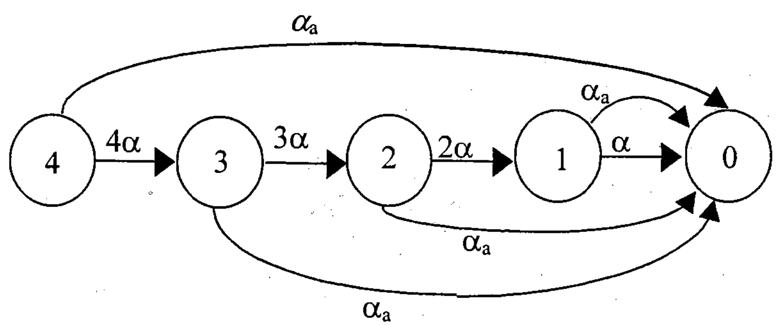


Figure A4. 2: Structure State Process for the Feed-Forward Make-to-Stock System

The make-to-stock system in consideration is failure prone. Let the observation period be [0, *t*]. Let the rv B represent the backorder level (which is the sum of all backorders at each operating end node) and I the inventory level of finished goods. Let O be the proportion of demand that is met immediately using the finished goods inventory. O is also known as the fill rate in the literature. Let x be the desired fill rate level and p be a specified probability. The questions of interest that can be answered using performability analysis are as follows.

For a given x and p, what would be a desired basestock level in order that $P(O > x) > p$?

Suppose that we can add a few more nodes to meet external demand, how many nodes need to be employed for a given S, x and p such that $P(O > x) > p$ is satisfied?

What are the values of E[B] and E[I]?

The above questions can be answered by computing over [0, t] the backorder-related, inventory-related, and fill-rate related performability distributions. As explained in Chapter 2, the performability rv is

$$Y_t(m) = \sum_{i=0}^{m} f_i \tau_i \tag{A4.1}$$

where τ_i is the total time [0, t] the structure state process (SSP) stays in state *i* and f_i is the reward associated with state *i*. The individual rewards, in our case fill-rate, average backorder level, or average inventory level are computed using the performance model. The sequential decomposition procedure can be used to find the measures for every structure state *i*. To find the distribution of $Y_t(m)$, we have to perform a transient analysis of the SSP.

Starting from state *m*, the evolution of the SSP during [0, t] can fall into three different categories.

- The SSP stays in state *m* throughout the interval, without making a transition to any other state.
- The SSP transits to state 0 directly from state *m* sometime during the interval and will therefore stay in state 0 for the rest of the interval.
- The SSP transits to state *m*-1 at some instant during the observation period. Its evolution during the rest of the interval will follow the same pattern as the original process, except that the initial state will be *m*-1.

This evolution of the SSP forms the basis for a recursive formulation over [0, t]. Let I denote an indicator function such that

I(C) = 1 if C is true

I(C) = 0 if C is false

Let p_{ij} ($i = 0, 1, ..., m$); $j \leq i$) denote the probability of a single-step transition for state i to state j. Note that, for $k = 1, 2, ..., m$,

$$p_{k,0} = \frac{\alpha_a}{\alpha_a + k\alpha} \tag{A4.2}$$

$$p_{k,k-1} = \frac{k\alpha}{\alpha_a + k\alpha} \tag{A4.3}$$

For $k = 0, 1, ..., m$, let $c_k = \alpha_a + k\alpha$. Since the failure distributions are assumed to exponential, the sojourn time in any state $k \geq 1$ is exponentially distributed with rate c_k. Therefore, $e^{-c_m t}$ gives the probability that the SSP stays in the initial state m throughout the interval [0, t]. To obtain the probability $P\{Y_t(m) < y\}$ where y > 0, the three observations presented earlier are used to obtain

$$\begin{aligned} P\{Y_t(m) < y\} = {} & I(f_m t < y)e^{-c_m t} \\ & + \int_0^t c_m e^{-c_m x} p_{m0} I([f_m x + f_0(t-x)] < y)dx \\ & + \int_0^t c_m e^{-c_m x} p_{m,m-1} P\{Y_{t-x}(m-1) < y - f_m x\}dx \end{aligned} \tag{A4.4}$$

The above equation gives a recursive formulation for computing the distribution of performability. Donatiello and Iyer (1987) present an efficient computational procedure to compute the performability distribution. The approach described above is what is available in the literature thus far in applying performability analysis to production systems. As long as the structure state process of the production or production-inventory system has a known solution approach and the performance analysis models are available, we can carryout the performability analysis in a straightforward manner. A new SSP would require the development of a solution approach for its transient analysis, which by itself is a challenging computational task.

The next section describes an alternative approach which was suggested by Prof. Y. Narahari. This approach involves the solution of a stochastic Petri net model.

A4.2 PERFORMABILITY ANALYSIS USING STOCHASTIC PETRI NETS

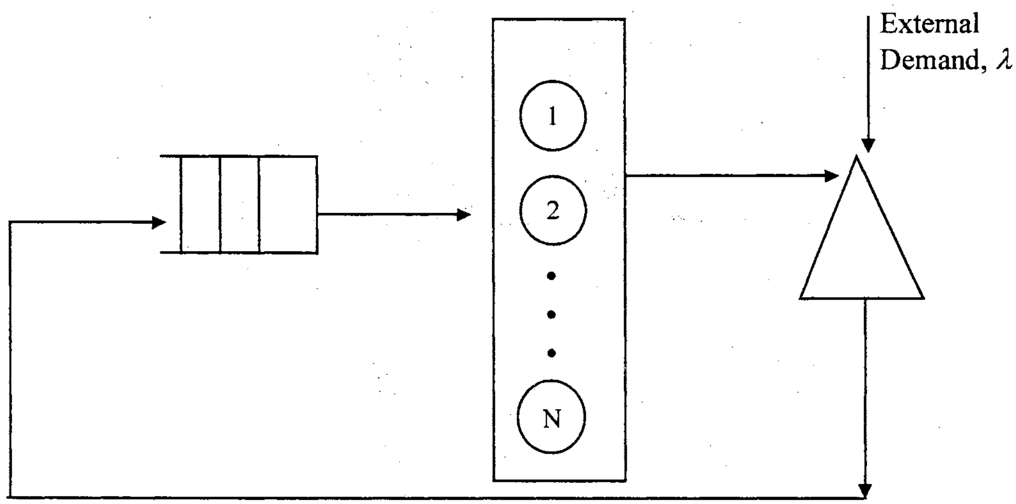


Figure A4. 3: A Single-Stage Make-to-Stock System with Multiple Servers

In this section, we use a single model to carryout the performability analysis. The system considered is a single-stage make-to-stock system with multiple servers as shown in Figure A4.3. N servers are operational at time t = 0. Demand is met from the output store. A one-for-one replenishment policy is assumed to be in practice at the output store. The part/order flow dynamics is similar to the systems described in earlier chapters.

The machines are prone to failures. When any one of the machines fails, the system continues to operate but at a slower rate since there is one less server available. It is assumed, that once a machine goes down, it is not repaired within the observation period.

Let the time to failure of each machine be exponentially distributed with rate α, and let the failures be independent of one another. Let the observation period be $[0, t]$, and B the backorder level and I the inventory level random variables. Examples of questions that can be answered using performability analysis are

- What are the values of E[B] and E[I]?
- What is average number of available machines?

We model this system using stochastic Petri nets. Petri nets, or place-transition nets, are classical models of concurrency, non-determinism, and control flow, first proposed in 1962 by Carl Adam Petri. They are bipartite graphs and provide an elegant and mathematically rigorous modeling framework for discrete event dynamical systems. The reader is referred to the book by Viswanadham and Narahari (1992) for an overview of Petri nets. A stochastic Petri net (SPN) is essentially a high-level model that generates a stochastic process. SPN-based performance evaluation basically consists of modeling the given system by an SPN, and automatically generating the stochastic process that governs the system behavior.

The Petri net model for the above system is shown in Figure A4.4.

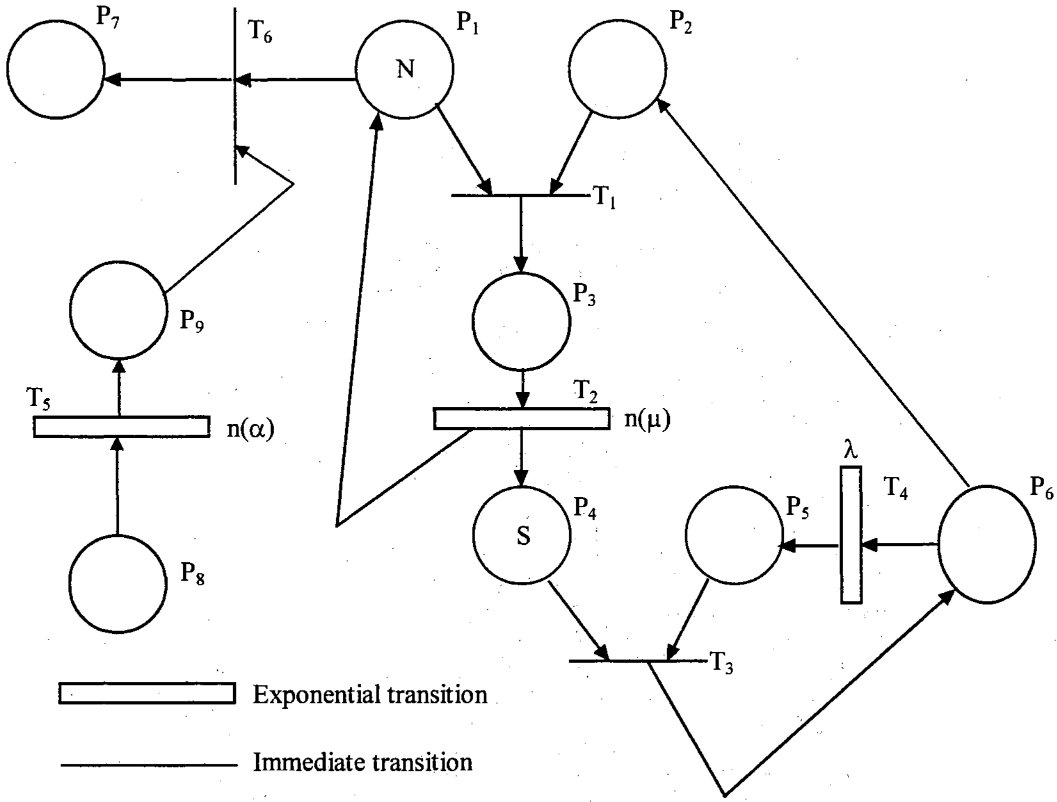


Figure A4. 4: Stochastic Petri Net Model of a Single-Stage Make-to-Stock System with Multiple Servers

The description of the place and transitions are given below:

Places:

P_1: Machine available for processing parts;

P_2: Parts waiting for machine;

P_3: Machine processing a part;

P_4: Finished parts available to meet demand;

P_5: External demand waiting for finished parts;

P_6: Generate external demand;

P_7: Failed machines;

P_8: Available machines; and

P_9: Failed machines waiting for completion of part.

Immediate (not timed) Transitions:

T_1: Machine begins processing a part;

T_3: External demand is satisfied; and

T_6: Machine fails after processing a part.

Exponential Transitions:

T_2: Processing of parts; the rate is dependent on number of tokens in P_3, rate = $M(P_3)\mu$;

T_4: Demand arrival into the system, rate = λ; and

T_5: Failing of machines, rate = $M(P_8)\ \alpha$.

A software tool called SPNP (Ciardo et al., 1989) was used to solve the above model. SPNP is one of the widely used Petri net tools developed by researchers at Duke University. Numerical results for some example cases are presented in Table A4.1. Arrival rate was 0.5, mean processing time for a part was 3.2 time units, and the mean time to failure was 240 time units. The observation period was set at 480 time units.

If it was a desired to have an average number of backorders of less than 15 units with the availability of at least 1 machine, a solution would be to have 9 parallel servers and a basestock level of 3 at the output store.

Table A4. 1: Performability Measures for a Single-Stage Make-to-Stock System with Multiple Servers

S	N	Average Backorders	Average Inventory	Average Number of Available Machines
3	3	26.840	0.050	0.411
12	3	25.812	0.536	0.411
3	6	20.558	0.206	0.819
12	6	18.413	2.020	0.819
3	9	14.777	0.404	1.220
12	9	11.873	3.694	1.220

N = number of machines; S = basestock level

A4.3 SUMMARY

In this appendix, we have shown how the technique of performability analysis can be useful in the analysis of the make-to-stock systems. This area of research is still a challenging one, because there is no general framework that can be used for any structure state process. Nevertheless, the technique is very useful in the design and analysis of not only make-to-stock systems but of systems which are fault-tolerant in general. The stochastic Petri net approach could be a viable alternative if the state space of the stochastic process generated is not very large.

VITA

Shankar Sivaramakrishnan

Candidate for the Degree of

Doctor of Philosophy

Thesis: ANALYTICAL MODELS FOR PERFORMANCE EVALUATION OF PRODUCTION-INVENTORY SYSTEMS

Major Field: Industrial Engineering & Management

Biographical:

Personal Data: Born in Ambattur, Madras, India, September, 10, 1967, the son of Sri S. Sivaramakrishnan and Smt. Vijayalakshmi Sivaraman

Education: Graduated from St. Mary's Matriculation Higher Secondary School in May, 1985; received the Master of Management Studies degree from Birla Institute of Technology & Science, Pilani, India in May, 1989; received the Master of Science degree from Oklahoma State University in May, 1993; completed requirements for the Doctor of Philosophy degree at Oklahoma State University in July, 1998.

Professional Experience: Research Associate, Sudaram Clayton Ltd. from January, 1989 to July, 1989; Research Assistant, School of Industrial Engineering and Management, Oklahoma State University from September, 1990 to July, 1992; Teaching Assistant, School of Industrial Engineering and Management, Oklahoma State University from August, 1992 to May, 1994; Research Associate, School of Industrial Engineering and Management, Oklahoma State University from August, 1993 to July, 1997; Research Scientist, PROS Strategic Solutions, Houston, Texas from August, 1997 to present.

Professional Membership: Alpha Pi Mu, IIE, INFORMS