TAIL ASYMPTOTICS OF QUEUEING NETWORKS WITH SUBEXPONENTIAL SERVICE TIMES

A Thesis
Presented to
The Academic Faculty

by

Jung-Kyung Kim

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology
August 2009

TAIL ASYMPTOTICS OF QUEUEING NETWORKS WITH SUBEXPONENTIAL SERVICE TIMES

Approved by:

Professor Hayriye Ayhan, Advisor
H. Milton Stewart School of Industrial
and Systems Engineering
Georgia Institute of Technology

Professor Robert D. Foley
H. Milton Stewart School of Industrial
and Systems Engineering
Georgia Institute of Technology

Professor David M. Goldsman
H. Milton Stewart School of Industrial
and Systems Engineering
Georgia Institute of Technology

Professor Bert Zwart
H. Milton Stewart School of Industrial
and Systems Engineering
Georgia Institute of Technology

Professor Joshua Reed
Leonard N. Stern School of Business
New York University

Date Approved: 1 July 2009

To myself and my parents...

ACKNOWLEDGEMENTS

Throughout my graduate studies, I am indebted to many people for their help and encouragement. First of all, I would like to thank my advisor Prof. Hayriye Ayhan. It is impossible for me to express in words my full appreciation. However, I would like to thank all of her advice and support during my graduate studies. I would not have completed this thesis without her infinite patience and understanding.

I would also like to thank all of my committee members, Prof. Robert Foley, Prof. David M. Goldsman, Prof. Bert Zwart, Prof. Yuri Bakhtin and Prof. Joshua Reed for their time and helpful suggestions which improved the quality of this thesis. Special thanks to Prof. David M. Goldsman and Prof. Bert Zwart for their letters of recommendation and to Prof. Joshua Reed for taking the time to fly to Atlanta for my defense.

I would also like to thank Prof. Seong-Hee Kim for her invaluable help and comments for assisting me in teaching courses and looking for a job as well as her recommendation letters.

I would like to thank all my friends who made my studies in USA enjoyable. I cannot list them all here, but I would like to especially mention my best friends, Jiheng Zhang, Eser Kırkızlar, Claudio Santiago, Dan Steffy, Myungcheol Doo and Woosuk Jung.

Finally, I would also like to thank my father Heedong Kim, my mother Heejung Goo, my brother Junghun Kim, and my sister-in-law Minam Goo for their love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This dissertation is concerned with the tail asymptotics of queueing networks with subexponential service time distributions. Our objective is to investigate the tail characteristics of key performance measures such as cycle times and waiting times on a variety of queueing models which may arise in many applications such as communication and manufacturing systems.

First, we focus on a general class of closed feedforward fork and join queueing networks under the assumption that the service time distribution of at least one station is subexponential. Our goal is to derive the tail asymptotics of transient cycle times and waiting times. Furthermore, we argue that under certain conditions the asymptotic tail distributions remain the same for stationary cycle times and waiting times. Finally, we provide numerical experiments in order to understand how fast the convergence of tail probabilities of cycle times and waiting times is to their asymptotic counter parts.

Next, we consider closed tandem queues with finite buffers between stations. We assume that at least one station has a subexponential service time distribution. We analyze this system under communication blocking and manufacturing blocking rules. We are interested in the tail asymptotics of transient cycle times and waiting times. Furthermore, we study under which conditions on system parameters a stationary regime exists and the transient results can be generalized to stationary counter parts. Finally, we provide numerical examples to understand the convergence behavior of the tail asymptotics of transient cycle times and waiting times.

Finally, we study open tandem queueing networks with subexponential service time distributions. We assume that number of customers in front of the first station

is infinite and there is infinite room for finished customers after the last station but the size of the buffer between two consecutive stations is finite. Using (max,+) linear recursions, we investigate the tail asymptotics of transient response times and waiting times under both communication blocking and manufacturing blocking schemes. We also discuss under which conditions these results can be generalized to the tail asymptotics of stationary response times and waiting times. Finally, we provide numerical examples to investigate the convergence of the tail probabilities of transient response times and waiting times to their asymptotic counter parts.

CHAPTER I

INTRODUCTION

Recent research has shown that in many queueing networks service times have subexponential distributions. For instance, in telecommunications setting, Fowler [35] argues that FTP (File Transfer Protocol) transfers have session sizes and session durations with subexponential distributions. Similar observations are made for the TELNET sessions in Paxson and Floyd [47] although TELNET is an application qualitatively quite different from FTP. Feldmann, Gilbert, Willinger and Kurtz [30] argue that these observations remain valid for today's World Wide Web (WWW) applications. Similarly, Arlitt and Williamson [2], Crovella and Bestavros [23] and Crovella and Lipsky [24] have shown evidence that the file sizes in Web have subexponential distributions.

This thesis is concerned with the tail characteristics of queueing networks with subexponential service times. Queueing networks are useful tools in modeling communication and manufacturing systems. Recent research has shown that subexponential distributions play a significant role in communication networks. However, such models are notoriously difficult to analyze since no closed form expressions exist for characteristics of these systems. Our objective is to analyze the tail asymptotic behavior of various performance measures such as cycle times (sojourn times) and waiting times on a variety of queueing models which may arise in many applications such as communication and manufacturing systems.

We consider three different systems with subexponential processing times in this thesis. The first system is a general class of closed feedforward fork and join queueing networks with subexponential service time distributions. Applications of fork and join

1

queues have been found in a variety of communication networks (such as data packets and computer processing systems) and manufacturing systems (such as assembly systems). We are interested in key performance measures such as cycle times and waiting times. In telecommunication systems with subexponential processing times, one is interested in the probability that these characteristics are bigger than a large value. For example, if W_n^k denotes the waiting time of the nth customer at node k in the system, one would like to get an expression for $P(W_n^k > x)$ as x gets large which is referred to as tail asymptotics. Examining the tail asymptotics of key performance measures is important in assessing how well a system is capable of preventing huge sojourn times and waiting times. Therefore, our objective is to derive expressions for the tail asymptotics of transient and stationary cycle times and waiting times. In order to characterize the transient cycle times and waiting times, we first define the notion of a path as a set of links in the opposite direction of customer flow and then provide upper and lower bounds for departure times of customers at the given station. In addition, we drive upper and lower bounds on transient cycle times and waiting times. Using these bounds, we obtain the tail asymptotics of transient cycle times and waiting times. Furthermore, we argue that under certain conditions on service times a stationary regime exists and the transient results can be generalized to stationary cycle times and waiting times. Finally, we provide numerical experiments in order to understand how fast the convergence of tail probabilities of cycle times and waiting times is to their asymptotic counter parts.

The second part of this thesis considers closed tandem queueing networks with finite buffers between stations. We assume that at least one station has a subexponential service time distribution. We analyze this system under the manufacturing blocking and communication blocking rules. More specifically, in the manufacturing blocking case, at the completion of service at station k, the customer can move to station $k+1$, if that buffer is not full. Otherwise it has to wait with server at station

k until the downstream buffer has a free space. On the other hand, in communication blocking, a server is not allowed to start service until space is available in the downstream buffer. These blocking mechanisms can appear in several applications; for example, window flow control in telecommunication systems and kanban blocking in manufacturing systems. Our objective is to derive expressions for the tail asymptotics of transient cycle times and waiting times. Furthermore, we study under which conditions on system parameters these tail asymptotics also hold for their stationary counter parts. Finally, we provide numerical examples to understand the convergence behavior of the tail asymptotics of cycle times and waiting times.

The final part of this thesis studies open tandem queues with subexponential service times and finite buffers. More specifically, we focus on K stations in tandem with an infinite number of customers in front of first station and infinite room for finished customers after last station. This model is operating under the manufacturing blocking and communication blocking rules. Baccelli, Schlegel, and Schmidt [16] and Dieker and Lelarge [27] have addressed the open tandem queues with finite buffers and provided the tail behavior of stationary response times. However, they assume stochastic input streams that are independent of the service process whereas we have an infinite supply of customers in front of the first station so that a new customer is accepted to the system as soon as the first server is free. Due to the explosive growth of the Internet and increasing demand for multimedia information on the web, transmission of multimedia over the Internet has received tremendous attention from academia and industry. Transmission of multimedia such as video and audio on the web could be modeled as our model. When video and audio are transported over the Internet to the receiver, video and audio data are first compressed and packetized at regular intervals and then saved in storage devices. After that each data packet is sent over the IP networks. We are interested in the tail behavior of transient and stationary response times and waiting times. Since a tandem queue with finite

3

buffers is an example of a (max,+) linear system, we use (max,+) linear recursions to obtain departure times of customers from the given station. Then, we calculate upper and lower bounds on the departure times. Also, we provide the upper and lower bounds on the transient response times and waiting times. Using these bounds, we compute the tail asymptotics of transient response times and waiting times. Also, we investigate whether there exist conditions on service times such that tail asymptotics for transient characteristics also hold for their stationary counter parts. Finally, we provide numerical examples to investigate the convergence of the tail probabilities of transient response times and waiting times to their asymptotic counter parts.

The remainder of this thesis is organized as follows. We provide a literature review on queues with subexponential service times in Chapter 2. Before stating the main results, we provide a brief description of a subexponential distribution and list its properties in Chapter 3. In Chapter 4, we provide the tail asymptotics of transient and stationary cycle times and waiting times for a closed fork and join queueing network with subexponential service times. We focus on cyclic queueing networks with finite buffers and provide the tail asymptotics of key performance measures such as cycle times and waiting times in Chapter 5. In Chapter 6, we study the tail characteristics for response times and waiting times in open tandem queues with finite buffers. Finally, we describe contributions of this research and our future directions in Chapter 7.

CHAPTER II

LITERATURE REVIEW

Queueing systems with subexponentiality arise in computer and communication systems. Some of the literature is interested in queueing networks with multiplexing on-off source with subexponential on periods; for example, Jelenkovíc and Lazar [39] or in queueing models with subexponential arrival streams; see, e.g., Jelenkovíc [38]. In this chapter, however, we are interested in stochastic queueing systems with subexponential service times. There is a vast body of literature on a variety of single queues with subexponential service times. They consider, for instance, multiserver queues, multiple arrival queues, Markov modulated G/G/1 queues, Generalized Processor Sharing (GPS) queues, and long-range dependent arrival queues with subexponential processing times. They investigate the tail characteristics of various performance measures such as waiting times, queue lengths, busy periods and sojourn times for these queues. Furthermore, in recent years, there has been some interest in extending the FIFO GI/GI/1 results to networks of queues (like tandem queues, split-match queues, generalized Jackson networks, (max,+) networks and so on).

This chapter is organized as follows. In Section 2.1, we provide a review of the literature on single queues with subexponential service times and in Section 2.2, we consider various networks under subexponential assumptions for service times.

2.1 *A Single Queue with Subexponential Service Times*

Performance impact of subexponential service times in single stage queues has been investigated extensively over the past decades. The study of a single-server queue with subexponential service times was first explored by Borovkov[18], Cohen[22], and Pakes [46]. In Pakes' paper, he focuses on the derivation of the tail asymptotics of

the waiting time in a FIFO GI/GI/1 queue under subexponentiality and therefore, he showed that when the residual service times are subexponential, the tail distribution of the service times will dominate the tail behavior of the stationary waiting times. Embrechts and Veraverbeke [29] compute the asymptotic behaviour of the probability of ruin function for the GI/G/1 queue with subexponential service times. In [54], Willekens and Teugels consider $M/G/1$ queues with FIFO service discipline and subexponential service times and they present asymptotic expansions for tail probabilities of the stationary waiting times. Moreover, they extend these results to the $M^{[X]}/G/1$ queue with batch arrivals.

Asmussen, Klüppelberg and Sigman [5] analyze the tail asymptotics of the steady-state queue length in GI/GI/1 queues with subexponential service times. Also, they have applications for queues with vacations and M/G/1 busy periods.

In [50], Scheller-Wolf and Sigman investigate the moments of the steady-state waiting time for FIFO GI/GI/s queues. Similarly, Whitt [53] focuses on a FIFO M/GI/s queue with unlimited waiting room and investigates the tail asymptotics of the stationary waiting time. For multiple Markovian arrival streams, Takine [52] provides subexponential asymptotics of the tail distribution of waiting times in stationary work-conserving single-server queues.

The result of Pakes [46] has later been generalized to Markov modulated G/G/1 queues by Jelenkovíc and Lazar [40]. Also, a similar Markov-modulated queueing model was studied by Asmussen, Henriksen and Klüppelberg [4]. Later, Asmussen [3] provides the asymptotic tail of the cycle maximum for the GI/G/1 actual waiting time process (which is a continuous time reflected Lévy process). In addition, Asmussen and Møller [6] consider bivariate regenerative Markov modulated queueing processes with subexponential increments. Tail asymptotics are obtained for both the maximum level over a regenerative cycle and the level itself when the increments of the level process have transition probabilities that are tail equivalent to a given fixed

subexponential distribution.

Borst, Boxma and Jelenković [19] analyze the behavior of long-tailed flows under the Generalized Processor Sharing (GPS) discipline. They focused on the exact workload asymptotics of an individual flow at a single node. Also, they show that for certain weight combinations an individual flow with long-tailed traffic characteristics is effectively served at a constant rate. The effective service rate may be interpreted as the maximum average traffic rate for the flow to be stable which is only influenced by the traffic characteristics of the other flows through their average rates.

In [56], Zwart characterizes the tail behaviour of the busy period distribution in the stable GI/G/1 queue in which the service time has a heavy (subexponential) tail. Later, Baltrūnas, Daley and Klüppelberg [17] extend the result to GI/GI/1 queue with subexponential service time distributions.

Asmussen, Schmidli and Schmidt [7] study short range dependent arrival process models. In a similar paper, Xia, Liu, Squillante and Zhang [55] provide asymptotic lower bounds for the tail distribution of the stationary waiting time under long-range dependent arrival process and i.i.d. subexponential service times.

Miyoshi [45] considers the model with a general stationary input rather than Markovian arrival stream input governed by a finite-state Markov chain and shows that the fundamental results still hold under some additional assumptions when the equilibrium residual service time distribution is subexponential.

Shang, Liu and Li [51] study the tail behavior of the stationary queue length of an M/G/1 retrial queue with subexponential service time distributions. Retrial queueing systems are characterized by the fact that any arriving customer who finds the server busy joins the retrial queue and retries for service in random order and at random interval. They show that the tail asymptotics of the stationary queue length in an M/G/1 retrial queue are determined by the tail of the stationary queue length in the corresponding standard M/G/1 queue.

Recently, for multi-server queues, Foss and Korshunov [32] investigate the asymptotic behavior of the distribution tail of the stationary waiting time in the GI/GI/2 queue with FCFS discipline and subexponential service times. Foss, Konstantopoulos and Zachary [31] study the asymptotic distribution of the maximum of a random walk, modulated by a regenerative process, when the increments have subexponential distributions. Here, "modulated" means that conditional on some background process with a regenerative structure, the random walk becomes a process with independent increments. They study the asymptotic behaviour of the maximum of the random walk in both discrete and continuous time. Boxma and Zwart [20] focus on the tail behavior of the response time of a job with subexponential service time distributions under various scheduling policies including preemptive and non-preemptive scheduling disciplines and discuss optimality properties. Denisov, Dieker and Shneer [25] investigate the distribution of the waiting time in a stable M/G/1 processor-sharing queue with traffic intensity $\rho < 1$ and Poisson arrivals of subexponential job sizes.

As the newest papers for subexponentiality, there are Ko and Tang [42], Foss and Richards [34], Geluk [36], Leipus and Šiaulys [44] and Foss, Korshunov and Zachary [33]. More specifically, Ko and Tang [42] and Foss and Richards [34] study the asymptotic tail probabilities of sums of dependent subexponential random variables. Geluk [36] investigates some closure properties for subexponential distributions. Leipus and Šiaulys [44] focus on the asymptotic behaviour of the finite-time ruin probability under subexponential claim sizes. Foss, Korshunov and Zachary [33] study convolutions of long-tailed and subexponential distributions. However, these all papers are beyond our scope.

2.2   *Networks with Subexponential Service Times*

Due to the rapid advances in computer and telecommunication systems, one needs to capture the complex situations that are observed in this area. Therefore, in the

last decade there has been a growing interest in queueing networks with subexponential service time distributions. However, there are not many existing results on the asymptotics of queueing networks with subexponential service times. The first papers in this area are provided by Baccelli, Schlegel and Schmidt [16] and Huang and Sigman [37]. They analyze open stochastic queueing networks with renewal arrivals and subexponential service time distributions. More specifically, Baccelli, Schlegel and Schmidt [16] deal with the tail behaviour of stationary response times in tandem networks of single server queues and then they extend the results to irreducible (max,+) linear systems. In a similar paper, Huang and Sigman [37] focus on the asymptotics of sojourn times and queue lengths in a variety of specific models including various tandem queues, split-match (fork-join) queues and feedforward generalized Jackson networks (GJN).

In [12], Baccelli and Foss extend these results to monotone-separable stochastic networks (which are networks whose state variables are homogeneous and monotone functions of the epochs of the arrival process. Some examples of this models are generalized Jackson networks, max-plus networks, polling systems, multiserver queues, and various classes of stochastic Petri nets). They provide upper and lower bounds for the tail asymptotics of the stationary maximal dater (which is the time to empty the network while stopping further arrivals. For instance, in a G/G/1 queue, this can be workload and in a FIFO tandem queue this can be end-to-end delay) in any network of this class. Furthermore, they obtain exact asymptotics for various special cases of these networks. Baccelli, Foss and Lelarge [13] provide the exact asymptotics of the tail of the stationary maximal dater in generalized Jackson networks of arbitrary topology with subexponential service times. However, they could not obtain the asymptotic behaviour of other state variables such as the stationary queue size in these networks. Baccelli, Lelarge and Foss [15] compute the exact tail asymptotics of stationary response times for both irreducible and reducible open stochastic event

9

graphs under the assumptions of renewal input and i.i.d. subexponential service times. In a recent paper, Dieker and Lelarge [27] study the tail asymptotics for functionals of the stationary solution of (max,+) linear recursions under subexponentiality assumptions in more complex networks; for example, the networks which have a FIFO event graph instead of a single server in each subnetwork. In addition, they apply the results to analyze the tail asymptotics of the resequencing delay. More specifically, packets have to be delivered to the destination in the order of transmission at the sender. However, due to the multi-path routing, packets may be misordered. Thus, networks need resequencing buffers for reordering. As a result, some of the packets have to wait in this buffers and they refer to this waiting time as resequencing delay. Kim and Ayhan [41] focus on open tandem networks with finite buffers and subexponential service times. They assume that number of customers in front of the first station is infinite and there is infinite room for finished customers after the last station. They provide the tail asymptotics of transient and stationary response times and waiting times. Details of Kim and Ayhan [41] are given in Chapter 6.

The above seven papers only concern open networks with subexponential service time distributions. To the best of our knowledge, there are two papers that study closed networks with subexponential processing times. Ayhan, Palmowski and Schlegel [9] investigate the tail distribution of transient and stationary cycle times and waiting times in closed tandem queues with subexponential service times. Ayhan and Kim [8] generalize the results of Ayhan, Palmowski and Schlegel [9] to a closed fork and join network. Details of Ayhan and Kim [8] are given in Chapter 4.

CHAPTER III

PRELIMINARIES

We provide the definition and some basic properties of subexponential distributions that will be needed in our analysis. The interested reader can refer to Embrechts, Klüppelberg and Mikosch [28] for a thorough study of these distributions. Moreover, the recent book Resnick [49] is a good reference for statistical analysis of these distributions.

Definition 3.0.1 *A distribution function F on $\mathbb{R}_+ = [0, \infty)$ with $F(x) < 1$ for all $x > 0$ is called subexponential $(F \in \mathcal{S})$ if*

$$\lim_{x \to \infty} \frac{\overline{F^{*n}}(x)}{\overline{F}(x)} = n,$$

where $\overline{F}(x) = 1 - F(x)$ and F^{*n} denotes the n-fold convolution of F with itself.

It can be shown that if the above condition holds for some $n \geq 2$, then it holds for all $n \geq 2$. One can immediately see that if $X_1, \ldots, X_n$ are independent random variables with distribution function $(F \in \mathcal{S})$ then

$$\lim_{x \to \infty} \frac{\mathbb{P}(X_1 + \cdots + X_n > x)}{\mathbb{P}(\max(X_1, \ldots, X_n) > x)} = 1.$$

In words, this means that the sum is likely to get large because one of the random variables gets large. It could be interpreted as a disaster in an insurance risk business or an unusually long processing time in a telecommunication network. The class $\mathcal{S}$ has some very useful properties. Those which are particularly used in this thesis are the following ones.

Lemma 3.0.1 *Let F and G be two distribution functions on $\mathbb{R}_+$ and assume that there exists a constant $c \in (0, \infty)$ with $\lim_{x \to \infty} \overline{G}(x)/\overline{F}(x) = c$. Then, $F \in \mathcal{S}$ if and only if $G \in \mathcal{S}$.*

Lemma 3.0.2 *Let F, G, and H be distribution functions on $\mathbb{R}_+$ such that $F \in \mathcal{S}$, $\lim_{x\to\infty} \overline{G}(x)/\overline{F}(x) = c_1$ and $\lim_{x\to\infty} \overline{H}(x)/\overline{F}(x) = c_2$, where $c_i \in [0, \infty)$ for $i = 1, 2$ and $c_1 + c_2 > 0$. Then,*

$$\lim_{x\to\infty} \frac{\overline{G * H}(x)}{\overline{F}(x)} = c_1 + c_2$$

where $*$ denotes convolution.

Lemma 3.0.3 *Let X and $Y \geq 0$ be independent random variables with distribution functions $F_X \in \mathcal{S}$ and F_Y, respectively. Then,*

$$\lim_{x\to\infty} \frac{\mathbb{P}(X - Y > x)}{\mathbb{P}(X > x)} = 1.$$

Lemma 3.0.4 *Let F and $G_1, ..., G_n$, $n \geq 1$, be distribution functions on $\mathbb{R}_+$ such that $\lim_{x\to\infty} \overline{G}_i(x)/\overline{F}(x) = c_i$ as $x \to \infty$; $c_i \geq 0$. Then,*

$$\lim_{x\to\infty} \frac{1 - \prod_{i=1}^{n} G_i(x)}{\overline{F}(x)} = \sum_{i=1}^{n} c_i.$$

Note that Lemma 3.0.4 holds for all distributions.

Corollary 3.0.1 *Let $F \in \mathcal{S}$ and let $F_1, ..., F_n$, $n \geq 1$, and $G_1, ..., G_m$, $m \geq 1$, be distribution functions on $\mathbb{R}_+$ such that $\lim_{x\to\infty} \overline{F}_i(x)/\overline{F}(x) = c_i$ with $c_i > 0$, $1 \leq i \leq n$, and $\overline{G}_i(x) = o(\overline{F}(x))$ for $1 \leq i \leq m$. Then,*

$$\lim_{x\to\infty} \frac{\overline{F_1 * ... * F_n * G_1 * ... * G_m}(x)}{\overline{F}(x)} = \sum_{i=1}^{n} c_i$$

where $*$ denotes convolution.

CHAPTER IV

A GENERAL CLASS OF CLOSED FORK AND JOIN
QUEUES WITH SUBEXPONENTIAL SERVICE TIMES

4.1 Introduction

In this chapter, we focus on a closed fork and join queueing network with subexponential service time distributions. Fork and join queues arise in many telecommunication and manufacturing applications (see Ko and Serfozo [43] for an excellent review of the literature on these networks). We consider a closed feedforward fork and join queueing network with deterministic routing and $K \geq 1$ stations. Such a network can be used to model a system where all customers depart the network from the same (root) station labelled K and as soon as a customer departs the network, new customers are accepted to the system through the entrance stations. A fork exists at each point that one or more customers can be initiated simultaneously. A join occurs whenever a customer is allowed to begin execution following the completion of one or more other customers. At the time of an arrival of a fork node, a customer is split into several customers which are served by each of the successive stations. At the time of an arrival of a join node, a customer has to wait for all other customers coming from preceding stations to complete the service. An example with $K = 7$ stations is shown in Figure 1. This system works in the following manner. All customers depart the network from the same station 7. As soon as a customer departs the network, a new customer is accepted to the system. That customer is split into three customers which will be sent to the entrance stations which are stations 1, 2, and 3. This process is corresponding to a fork. Station 4 needs a customer coming from station 1 and coming from station 2. Those customers join and receive the service at station 4.

This process is corresponding to a join. Note that the customers are assumed to be distinguishable. Let $p(i)$ be the set of immediate predecessors of station i. Similarly, $p(A)$ denotes the set of immediate predecessors of set $A \subseteq \{1, \ldots, K\}$. Since the network is feedforward, we label the remaining $K - 1$ stations such that if station $j \in p(i)$ then $j < i$ unless $p(i) = \{K\}$ (in which case $K > i$). Furthermore, we assume that if $K \in p(i)$ for some $i \in \{1, \ldots, K - 1\}$, then $p(i) = \{K\}$. Therefore, the entrance nodes have no other predecessors other than station K. We use I to denote the set of entrance stations. Thus, $i \in I$ if $p(i) = \{K\}$. Moreover, $p^n(i)$ for $n \geq 1$ denotes $\underbrace{p(p(\ldots (p(i)) \ldots))}_{n}$. The following notion of a *path* will be used in our developments.

Definition 4.1.1 *If $j \in p^k(i)$ for some $k \geq 1$, we define a path α_{ij} from station i to station j as a set of links $\{(i, i_1), (i_1, i_2), \ldots, (i_{k-1}, j)\}$ such that $i_1, i_2, \ldots, i_{k-1} \in \{1, \ldots, K\} \setminus \{j\}$, $i_1 \in p(i)$, $i_r \in p(i_{r-1})$, $r = 2, \ldots, k-1$, $j \in p(i_{k-1})$, and if $i \in I$, none of $i_1, i_2, \ldots, i_{k-1}$ can belong to I and if $i \notin I$, at most one of $i_1, i_2, \ldots, i_{k-1}$ can belong to I. Moreover, for $i \notin I$, if $j \in p^k(i)$ and $j \in p^l(i)$ for some $k, l \geq 1$ and if there exist $\{(i, i_1), (i_1, i_2), \ldots, (i_{k-1}, j)\}$ such that $i_1 \in p(i)$, $i_r \in p(i_{r-1})$, $r = 2, \ldots, k-1$, $j \in p(i_{k-1})$, with none of $i_1, \ldots, i_{k-1}$ belonging to I and $\{(i, i'_1), (i'_1, i'_2), \ldots, (i'_{l-1}, j)\}$ such that $i'_1 \in p(i)$, $i'_r \in p(i_{r-1})$, $r = 2, \ldots, l-1$, $j \in p(i'_{l-1})$, with one of $i'_1, \ldots, i'_{l-1}$ belonging to I, then $\{(i, i'_1), (i'_1, i'_2), \ldots, (i'_{l-1}, j)\}$ cannot form a path from station i to station j.*

Thus, a path is traversed in the opposite direction of customer flow. Note that there could be more than one path from station i to station j. For convenience, let $S^{\alpha_{ij}}$ denote the set of stations visited along the path α_{ij}. Hence, in the above definition, if $i \neq j$, then $S^{\alpha_{ij}} = \{i, i_1, i_2, \ldots, i_{k-1}, j\}$ otherwise $S^{\alpha_{ij}} = \{i, i_1, i_2, \ldots, i_{k-1}\}$. We use $N_{i,j}$ to denote the initial number of customers in front of station i coming from
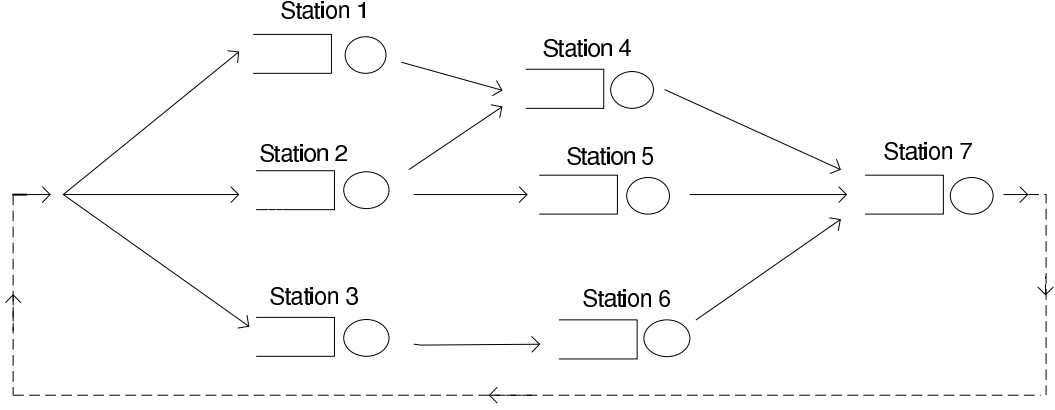
Figure 1: A closed feedforward fork and join queueing network with $K = 7$ stations.

station $j \in p(i)$ and assume that for all $i \in I$

$$\sum_{(k,l)\in\alpha_{ii}} N_{k,l} = N \geq 1, \ \forall \ \alpha_{ii}. \tag{1}$$

That is if $j \in p(i)$ and $|p(i)| > 1$ (where $|A|$ denotes the cardinality of set A), then $N_{i,j}$ is known for all $j \in p(i)$. Note that N is the total number of customers in the system. We assume that the network is deadlock-free which implies that the system will not evolve into the situation that where one or a set of the stations can never start processing customers (see for example pages 60–61 of Baccelli, Cohen, Olsder and Quadrat [11] for a formal definition of deadlock-free). Service times at station $k \in \{1, \ldots, K\}$ are independent and identically distributed random variables $\{B_n^k\}$ with distribution function $B_k(\cdot)$. The sequence of service times at each station is independent of the service times at the other stations. Moreover, we assume that there exists a subexponential distribution $F(\cdot)$ ($F \in \mathcal{S}$) and there exist constants $c_k \in [0, \infty)$ with $\sum_{k=1}^{K} c_k > 0$ such that for all $k \in \{1, \ldots, K\}$

$$\lim_{x\to\infty} \frac{\overline{B}_k(x)}{\overline{F}(x)} = c_k \tag{2}$$

where $\overline{F}(x) = 1 - F(x)$. For this network, we are interested in the tail behavior of transient and stationary cycle times (time between the successive departures of the same customer from a given station) and waiting times at each station.

15

Rest of this chapter is organized as follows. In Section 4.2, we provide some preliminary results. Section 4.3 focuses on the tail asymptotics of transient and stationary cycle times and waiting times. Section 4.4 provides numerical results.

4.2 Preliminaries

One can immediately see that (1) implies that for all $j \in \{1, \ldots, K\}$

$$\sum_{(k,l) \in \alpha_{jj}} N_{k,l} = N \geq 1, \; \forall \; \alpha_{jj}. \tag{3}$$

This follows from the observation that each α_{jj} contains the same links as some α_{ii} for $i \in I$. The next result says that total number of initial customers along any path from station i to station j, $i, j \in \{1, \ldots, K\}$ is the same.

Lemma 4.2.1 *For all* α_{ij}, $i, j \in \{1, \ldots, K\}$, $\sum_{(u,v) \in \alpha_{ij}} N_{u,v}$ *is equal to the same value.*

Proof We assume that $i \neq j$ since the case $i = j$ is given in (3). First suppose that $i \notin I$ and α_{ij} consists of links $\{(i, i_1), (i_1, i_2), \ldots, (i_{k-1}, i_k), (i_k, j)\}$ such that $i_1, i_2, \ldots, i_k \notin I$. Then it follows from our definition of a path that all α'_{ij} should have the same structure. That is, if α'_{ij} is composed of links $\{(i, i'_1), (i'_1, i'_2), \ldots, (i'_{l-1}, i'_l), (i'_l, j)\}$ then $i'_1, i'_2, \ldots, i'_l \notin I$. Now add the links $\{(j, j_1), (j_1, j_2), \ldots, (j_m, i)\}$ to α_{ij} and α'_{ij} so that one obtains the paths α_{ii} and α'_{ii}. Note that this is possible since from any station j we have a path to one of the stations that belong to I. Then

$$\begin{aligned}
\sum_{(u,v) \in \alpha_{ij}} N_{u,v} &= \sum_{(u,v) \in \alpha_{ii}} N_{u,v} - (N_{j,j_1} + N_{j_1,j_2} + \ldots + N_{j_m,i}) \\
&= \sum_{(u,v) \in \alpha'_{ii}} N_{u,v} - (N_{j,j_1} + N_{j_1,j_2} + \ldots + N_{j_m,i}) \\
&= \sum_{(u,v) \in \alpha'_{ij}} N_{u,v}
\end{aligned}$$

where the second equality follows from (3). Next assume that $i \notin I$ and α_{ij} consists of links $\{(i, i_1), (i_1, i_2), \ldots, (i_l, K), (K, i_{l+2}), \ldots, (i_k, j)\}$ for some $k, l \geq 1$, where $i_l \in I$.

Then if there are more than one path from station i to station j, there exits α'_{ij} which consists of the links $\{(i, i_1), (i_1, i_2),\ldots, (i_l, K), (K, i'_{l+2}),\ldots, (i'_{k'}, j)\}$ for some $k' \geq 1$ and/or α''_{ij} which consists of the links $\{(i, i'_1), (i'_1, i'_2), \ldots, (i'_{l'}, K), (K, i_{l+2}),\ldots,(i_k, j)\}$ for some $l' \geq 1$, where $i'_{l'} \in I$. Then for $r \geq 1$,

$$
\begin{aligned}
\sum_{(u,v)\in\alpha_{ij}} N_{u,v} &= (N_{i,i_1} + N_{i_1,i_2} + \ldots + N_{i_l,K}) + (N_{K,i_{l+2}} + \ldots + N_{i_k,j} + N_{j,j_1} + \ldots + N_{j_r,K}) \\
&\quad -(N_{j,j_1} + \ldots + N_{j_r,K}) \\
&= (N_{i,i_1} + N_{i_1,i_2} + \ldots + N_{i_l,K}) + (N_{K,i'_{l+2}} + \ldots + N_{i'_{k'},j} + N_{j,j_1} + \ldots + N_{j_r,K}) \\
&\quad -(N_{j,j_1} + \ldots + N_{j_r,K}) \\
&= \sum_{(u,v)\in\alpha'_{ij}} N_{u,v}
\end{aligned}
$$

where $\{(K, i_{l+2}),\ldots, (i_k, j), (j, j_1),\ldots, (j_r, K)\}$ and $\{(K, i'_{l+2}),\ldots, (i'_{k'}, j), (j, j_1),\ldots, (j_r, K)\}$, for some $r \geq 1$, are sets of links forming paths from station K to itself and the second equality holds since $\sum_{(u,v)\in\alpha_{KK}} N_{u,v} = N$ for all α_{KK}. Similarly, for $n \geq 1$,

$$
\begin{aligned}
\sum_{(u,v)\in\alpha_{ij}} N_{u,v} &= (N_{i,i_1} + N_{i_1,i_2} + \ldots + N_{i_l,K} + N_{K,j_1} + \ldots + N_{j_n,i}) \\
&\quad -(N_{K,j_1} + \ldots + +N_{j_n,i}) + N_{K,i_{l+2}} + \ldots + N_{i_k,j} \\
&= (N_{i,i'_1} + N_{i'_1,i'_2} + \ldots + N_{i'_{l'},K} + N_{K,j_1} + \ldots + N_{j_n,i}) \\
&\quad -(N_{K,j_1} + \ldots + +N_{j_n,i}) + N_{K,i_{l+2}} + \ldots + N_{i_k,j} \\
&= \sum_{(u,v)\in\alpha''_{ij}} N_{u,v}
\end{aligned}
$$

where $\{(i, i_1),\ldots, (i_l, K), (K, j_1),\ldots, (j_n, i)\}$ and $\{(i, i'_1),\ldots, (i'_{l'}, K), (K, j_1),\ldots, (j_n, i)\}$, for some $n \geq 1$, are sets of links forming paths from station i to itself and the second equality holds since $\sum_{(u,v)\in\alpha_{ii}} N_{u,v} = N$ for all α_{ii}. Thus, $\sum_{(u,v)\in\alpha'_{ij}} N_{u,v} = \sum_{(u,v)\in\alpha_{ij}} N_{u,v} = \sum_{(u,v)\in\alpha''_{ij}} N_{u,v}$

Note that the proof of the case $i \in I$ is the same as showing $\sum_{(u,v)\in\alpha'_{ij}} N_{u,v} = \sum_{(u,v)\in\alpha_{ij}} N_{u,v}$ when we set $i_l = i$. $\qquad\square$

For notational convenience, define

$$\mathcal{N}_{k,j} = \sum_{(u,v) \in \alpha_{kj}} N_{u,v}$$

as the total number of initial customers along any path from station k to station j for $k, j \in \{1, \ldots, K\}$.

We next provide upper and lower bounds on the departure times. Let X_n^k denote the departure time of the nth customer from station $k \in \{1, \ldots, K\}$. We have

$$X_n^k = \max\{B_n^k + X_{n-1}^k, \max_{j \in p(k)} (B_n^k + X_{n-\mathcal{N}_{k,j}}^j)\} \tag{4}$$

where $X_n^k = 0$ and $B_n^k = 0$ for all $n \leq 0$ and $k \in \{1, \ldots, K\}$. The following proposition provides an upper bound on X_n^k.

Proposition 4.2.1 *For all $k \in \{1, \ldots, K\}$ and $n \geq 1$,*

$$X_n^k \leq \sum_{\substack{j=1 \\ j \neq k}}^{K} \sum_{r=1}^{n-\mathcal{N}_{k,j}} B_r^j + \sum_{r=1}^{n} B_r^k$$

with the convention that the summation over an empty set is zero.

Proof Note that the nth customer served at station k is the $(n - \mathcal{N}_{k,j})^{\text{th}}$ customer served at station $j \neq k$ and if these customers were served sequentially at the stations, one would obtain the above upper bound. $\qquad \square$

One can also obtain the following lower bound on X_n^k from the observation that the nth customer served at station k is the $(n - \mathcal{N}_{k,j})^{\text{th}}$ customer served at station $j \neq k$.

Proposition 4.2.2 *For all $k \in \{1, \ldots, K\}$ and $n \geq 1$,*

$$X_n^k \geq \max\{\max_{\substack{j=1,\ldots,K \\ j \neq k}} \sum_{r=1}^{n-\mathcal{N}_{k,j}} B_r^j, \sum_{r=1}^{n} B_r^k\}$$

with the convention that the summation over an empty set is equal to zero and the maximization over an empty set is equal to $-\infty$.

18

4.3 Cycle Times and Waiting Times

In this section, we first provide the tail asymptotics of transient cycle times and waiting times and then argue that the asymptotic tail distribution remains the same for stationary cycle times and waiting times under certain assumptions.

4.3.1 Transient Characteristics

Let C_n^k denote the nth cycle time at station $k \in \{1, \ldots, K\}$. By a cycle time, we mean the time between the successive departures of the same customer from a given station. Then it follows from (3) that

$$C_n^k = X_{n+N}^k - X_n^k.$$

The next proposition provides the tail asymptotics for C_n^k for $k \in \{1, \ldots, K\}$.

Proposition 4.3.1 *For all $k \in \{1, \ldots, K\}$ and for all $n \geq \max_{\substack{j=1,\ldots,K \\ j \neq k}} \mathcal{N}_{k,j}$,*

$$\lim_{x \to \infty} \frac{\mathbb{P}\{C_n^k > x\}}{\overline{F}(x)} = N \sum_{j=1}^{K} c_j$$

where the convergence is uniform in n.

Proof To obtain an upper bound for C_n^k, we define the set of service times $\mathcal{C}_n^k$ as

$$\mathcal{C}_n^k = \bigcup_{\substack{j=1 \\ j \neq k}}^{K} \bigcup_{r=n+1-\mathcal{N}_{k,j}}^{n+N-\mathcal{N}_{k,j}} \{B_r^j\} \cup \bigcup_{r=n+1}^{n+N} \{B_r^k\}.$$

Note that at least one of the service times in $\mathcal{C}_n^k$ must be in progress at any time in the interval $[X_n^k, X_{n+N}^k]$ and there is no other service time (other than those in $\mathcal{C}_n^k$) that could take place in this time interval and could have an effect on C_n^k, and then we have

$$C_n^k \leq \sum_{\substack{j=1 \\ j \neq k}}^{K} \sum_{r=n+1-\mathcal{N}_{k,j}}^{n+N-\mathcal{N}_{k,j}} B_r^j + \sum_{r=n+1}^{n+N} B_r^k.$$

19

From Corollary 3.0.1, we obtain that for all $n \geq \max_{\substack{j=1,\ldots,K \\ j \neq k}} \mathcal{N}_{k,j}$,

$$
\limsup_{x \to \infty} \frac{\mathbb{P}\{C_n^k > x\}}{\overline{F}(x)} \leq \limsup_{x \to \infty} \frac{\mathbb{P}\{\sum_{j=1}^{K} \sum_{r=1}^{N} B_r^j > x\}}{\overline{F}(x)}
$$

$$
= N \sum_{j=1}^{K} c_j. \tag{5}
$$

Now we derive a lower bound on C_n^k. One can observe that all service times in $\mathcal{C}_n^k$ take place within the time interval $[X_{n-N}^k, X_{n+N}^k]$ and that the service times that occur in the interval from X_{n-N}^k to X_n^k do not have an effect on C_n^k. Therefore,

$$
C_n^k \geq \max\{\max_{\substack{j=1,\ldots,K \\ j \neq k}} \max_{r=n+1-\mathcal{N}_{k,j},\ldots,n+N-\mathcal{N}_{k,j}} B_r^j, \max_{r=n+1,\ldots,n+N} B_r^k\} - C_{n-N}^k
$$

$$
\geq \max\{\max_{\substack{j=1,\ldots,K \\ j \neq k}} \max_{r=n+1-\mathcal{N}_{k,j},\ldots,n+N-\mathcal{N}_{k,j}} B_r^j, \max_{r=n+1,\ldots,n+N} B_r^k\}
$$

$$
-\{\sum_{\substack{j=1 \\ j \neq k}}^{K} \sum_{r=n-N+1-\mathcal{N}_{k,j}}^{n-\mathcal{N}_{k,j}} B_r^j + \sum_{r=n-N+1}^{n} B_r^k\}
$$

where the second inequality follows the upper bound on C_{n-N}^k. Since the upper bound on C_{n-N}^k is independent of the service times in set $\mathcal{C}_n^k$, from Lemmas 3.0.3, 3.0.4 and Corollary 3.0.1, for $n \geq \max_{\substack{j=1,\ldots,K \\ j \neq k}} \mathcal{N}_{k,j}$, we have

$$
\liminf_{x \to \infty} \frac{\mathbb{P}\{C_n^k > x\}}{\overline{F}(x)} \geq \liminf_{x \to \infty} \frac{\mathbb{P}\{\max_{j=1,\ldots,K} \max_{r=1,\ldots,N} B_r^j\}}{\overline{F}(x)}
$$

$$
= N \sum_{j=1}^{K} c_j
$$

which together with (5) completes the proof for the asymptotics of the nth cycle time at station k. □

Remark 4.3.1 As in Ayhan, Palmowski, and Schlegel [9], the tail asymptotics of cycle times at station k for all $k \in \{1, \ldots, K\}$ is proportional to the product of the number of customers in the network and the sum of the constants c_j, $j = 1, \ldots, K$. Thus, the asymptotic tail behaviour of the cycle times has the same structure for cyclic tandem queues and fork and join queues.

In the next proposition, we provide the tail asymptotics of transient waiting times. Let W_n^k be the waiting time of the nth customer until the start of his service at station k with $|p(k)| = 1$. Similarly, let $W_n^{k,l}$ be the waiting time of the nth arriving customer from station $l \in p(k)$ at station k with $|p(k)| > 1$. Then, if $|p(k)| = 1$, for $n \geq 1$,

$$W_n^k = \max\{X_{n-1}^k - X_{n-N_{k,p(k)}}^{p(k)}, 0\}$$

and if $|p(k)| > 1$, for $l \in p(k)$ and $n \geq 1$,

$$W_n^{k,l} = \max\{X_{n-1}^k - X_{n-N_{k,l}}^l, \max_{\substack{j \in p(k) \\ j \neq l}} X_{n-N_{k,j}}^j - X_{n-N_{k,l}}^l, 0\}.$$

For a station $k \in \{1, \ldots, K\}$, define $I^k \subseteq I$ such that if $j \in I^k$, then $K \notin S^{\alpha_{kj}}$ for all α_{kj} unless $k = K$ in which case $I^k = I$. Let $S^k = \bigcup_{j \in I^k} \bigcup_{\alpha_{kj}} S^{\alpha_{kj}}$. Note that if $k \in I$, then $I^k = \emptyset$ and hence, $S^k = \emptyset$. Finally, for k with $|p(k)| > 1$ and $l \in p(k)$, define $A^{k,l} = S^k \setminus \{S^l \bigcup \{k, l\}\}$. Thus, $A^{k,l}$ is the set of stations that a customer can possibly visit until he reaches station k given that he does not go through station l.

Proposition 4.3.2 *For $k \in \{1, \ldots, K\}$ with $|p(k)| = 1$ and $B_k \in \mathcal{S}$ for all $n \geq N$*

$$\lim_{x \to \infty} \frac{\mathbb{P}\{W_n^k > x\}}{\overline{F}(x)} = (N-1)c_k$$

uniformly in n and for $k \in \{1, \ldots, K\}$ with $|p(k)| > 1$ and $l \in p(k)$, if $B_k \in \mathcal{S}$ or $B_j \in \mathcal{S}$ for some $j \in A^{k,l}$ then for all $n \geq \max_{j \in A^{k,l}}\{N + \mathcal{N}_{k,j}\}$

$$\lim_{x \to \infty} \frac{\mathbb{P}\{W_n^{k,l} > x\}}{\overline{F}(x)} = (N-1)c_k + N \sum_{j \in A^{k,l}} c_j$$

uniformly in n.

Proof For the case with $|p(k)| = 1$, we derive an upper bound on W_n^k. Let

$$\mathcal{W}_n^k = \bigcup_{r=n-N+1}^{n-1} \{B_r^k\}.$$

Note that at least one of the service times in $\mathcal{W}_n^k$ must be in progress at anytime in the time interval $[X_{n-N_{k,j}}^j, X_{n-1}^k]$ if $X_{n-1}^k \geq X_{n-N_{k,j}}^j$ (otherwise $W_n^k = 0$) where

21

$p(k) = \{j\}$ and there is no other service time (other than those in $\mathcal{W}_n^k$) that could take place in this interval and could affect W_n^k. Then, we have

$$W_n^k \leq \sum_{r=n-N+1}^{n-1} B_r^k.$$

Therefore, from Corollary 3.0.1, for all $n \geq N$

$$
\begin{aligned}
\limsup_{x\to\infty} \frac{\mathbb{P}\{W_n^k > x\}}{\overline{F}(x)} &\leq \limsup_{x\to\infty} \frac{\mathbb{P}\{\sum_{r=1}^{N-1} B_r^k > x\}}{\overline{F}(x)} \\
&= (N-1)c_k.
\end{aligned}
\tag{6}
$$

For a lower bound on W_n^k, we observe that all service times in $\mathcal{W}_n^k$ take place in the time interval $[X_{n-N}^k, X_{n-1}^k]$. Moreover, completed service times that take place in the interval from X_{n-N}^k to $X_{n-N_{k,j}}^j$ do not have an effect on W_n^k. Since $X_{n-N}^k \geq X_{n-N-N_{k,j}}^j$, then we obtain

$$
\begin{aligned}
W_n^k &\geq \max_{r=n-N+1,\dots,n-1} B_r^k - C_{n-N-N_{k,j}}^j \\
&\geq \max_{r=n-N+1,\dots,n-1} B_r^k - \{\sum_{\substack{m=1\\m\neq j}}^{K} \sum_{r=n-N-N_{k,j}+1-\mathcal{N}_{j,m}}^{n-N_{k,j}-\mathcal{N}_{j,m}} B_r^k + \sum_{r=n-N-N_{k,j}+1}^{n-N_{k,j}} B_r^j\}
\end{aligned}
$$

where the upper bound on $C_{n-N-N_{k,j}}^j$ follows from the proof of Proposition 4.3.1. Since the upper bound on $C_{n-N-N_{k,j}}^j$ is independent of the service times in $\mathcal{W}_n^k$, from Lemma 3.0.3 and 3.0.4, for all $n \geq N$, we have

$$
\begin{aligned}
\liminf_{x\to\infty} \frac{\mathbb{P}\{W_n^k > x\}}{\overline{F}(x)} &\geq \liminf_{x\to\infty} \frac{\mathbb{P}\{\max_{r=1,\dots,N-1} B_r^k\}}{\overline{F}(x)} \\
&= (N-1)c_k
\end{aligned}
$$

and this together with (6) completes the proof.

Next, we provide the proof for the tail asymptotics of $W_n^{k,l}$ when $|p(k)| > 1$ and $l \in p(k)$. Let

$$\mathcal{W}_n^{k,l} = \bigcup_{j\in A^{k,l}} \bigcup_{r=n+1-N-\mathcal{N}_{k,j}}^{n-\mathcal{N}_{k,j}} \{B_r^j\} \cup \bigcup_{r=n+1-N}^{n-1} \{B_r^k\}.$$

22

Note that if $\max\{\max_{\substack{j\in p(k)\\j\neq l}} X^j_{n-N_{k,j}}, X^k_{n-1}\} > X^l_{n-N_{k,l}}$ (otherwise $W^{k,l}_n = 0$), then at least one of the service times in $\mathcal{W}^{k,l}_n$ must be in progress in the time interval $\big[X^l_{n-N_{k,l}}, \max\{X^k_{n-1}, \max_{\substack{j\in p(k)\\j\neq l}} X^j_{n-N_{k,j}}\}\big]$. Moreover, there is no other service time (other than those in $\mathcal{W}^{k,l}_n$) that could take place in this time interval and have an effect on $W^{k,l}_n$. Thus,

$$W^{k,l}_n \leq \sum_{j\in A^{k,l}} \sum_{r=n+1-N-\mathcal{N}_{k,j}}^{n-\mathcal{N}_{k,j}} B^j_r + \sum_{r=n+1-N}^{n-1} B^k_r$$

and from Corollary 3.0.1, for all $n \geq \max_{j\in A^{k,l}}\{N + \mathcal{N}_{k,j}\}$

$$\begin{aligned}
&\limsup_{x\to\infty} \frac{\mathbb{P}\{W^{k,l}_n > x\}}{\overline{F}(x)}\\
\leq\; &\limsup_{x\to\infty} \frac{\mathbb{P}\{\sum_{j\in A^{k,l}}\sum_{r=1}^N B^j_r + \sum_{r=1}^{N-1} B^k_r > x\}}{\overline{F}(x)}\\
=\; &(N-1)c_k + N\sum_{j\in A^{k,l}} c_j.
\end{aligned} \tag{7}$$

Note that all the service times in $\mathcal{W}^{k,l}_n$ take place in the interval

$$\big[\min\{X^k_{n-N}, \min_{j\in A^{k,l}} X^j_{n-N-\mathcal{N}_{k,j}}\}, \max\{X^k_{n-1}, \max_{\substack{j\in p(k)\\j\neq l}} X^j_{n-N_{k,j}}\}\big]$$

and the completed service times that take place in the time interval from

$$\min\{X^k_{n-N}, \min_{j\in A^{k,l}} X^j_{n-N-\mathcal{N}_{k,j}}\} \quad\text{to}\quad X^l_{n-N_{k,l}}$$

do not have an effect on $W^{k,l}_n$. We have

$$\begin{aligned}
&X^l_{n-N_{k,l}} - \min\{X^k_{n-N}, \min_{j\in A^{k,l}} X^j_{n-N-\mathcal{N}_{k,j}}\}\\
=\; &\max\{X^l_{n-N_{k,l}} - X^k_{n-N}, \max_{j\in A^{k,l}}\{X^l_{n-N_{k,l}} - X^j_{n-N-\mathcal{N}_{k,j}}\}\}\\
\leq\; &\max\{X^l_{n-N_{k,l}} - X^l_{n-N-N_{k,l}}, \max_{j\in A^{k,l}}\{X^l_{n-N_{k,l}} - X^l_{n-2N-N_{k,l}}\}\}\\
=\; &X^l_{n-N_{k,l}} - X^l_{n-2N-N_{k,l}}\\
=\; &C^l_{n-N-N_{k,l}} + C^l_{n-2N-N_{k,l}}
\end{aligned}$$

where the inequality follows from (4). Then

$$
\begin{aligned}
W_n^{k,l} \;\geq\;& \max\{\max_{j\in A^{k,l}} \max_{r=n+1-N-\mathcal{N}_{k,j},\dots,n-\mathcal{N}_{k,j}} B_r^j, \max_{r=n+1-N,\dots,n-1} B_r^k\} \\
& -C_{n-N-N_{k,l}}^l - C_{n-2N-N_{k,l}}^l \\[2mm]
\geq\;& \max\{\max_{j\in A^{k,l}} \max_{r=n+1-N-\mathcal{N}_{k,j},\dots,n-\mathcal{N}_{k,j}} B_r^j, \max_{r=n+1-N,\dots,n-1} B_r^k\} \\
& -\sum_{\substack{j=1\\ j\neq l}}^{M} \sum_{r=n-N-N_{k,l}+1-\mathcal{N}_{l,j}}^{n-N_{k,l}-\mathcal{N}_{l,j}} B_r^j - \sum_{r=n-N-N_{k,l}+1}^{n-N_{k,l}} B_r^l \\
& -\sum_{\substack{j=1\\ j\neq l}}^{M} \sum_{r=n-2N-N_{k,l}+1-\mathcal{N}_{l,j}}^{n-N-N_{k,l}-\mathcal{N}_{l,j}} B_r^j - \sum_{r=n-2N-N_{k,l}+1}^{n-N-N_{k,l}} B_r^l
\end{aligned}
$$

where the upper bounds on $C_{n-N-N_{k,l}}^l$ and $C_{n-2N-N_{k,l}}^l$ are obtained by summing the service times that belong to the sets $\mathcal{C}_{n-N-N_{k,l}}^l$ and $\mathcal{C}_{n-2N-N_{k,l}}^l$, respectively (see the proof of Proposition 4.3.1). Since the service times in $\mathcal{W}_n^{k,l}$ are independent of the service times in $\mathcal{C}_{n-N-N_{k,l}}^l$ and $\mathcal{C}_{n-2N-N_{k,l}}^l$, from Lemmas 3.0.3 and 3.0.4, for all $n \geq \max_{j\in A^{k,l}}\{N + \mathcal{N}_{k,j}\}$

$$
\begin{aligned}
& \liminf_{x\to\infty} \frac{\mathbb{P}\{W_n^{k,l} > x\}}{\overline{F}(x)} \\
\geq\;& \liminf_{x\to\infty} \frac{\mathbb{P}\{\max\{\max_{j\in A^{k,l}} \max_{r=1,\dots,N} B_r^j, \max_{r=1,\dots,N-1} B_r^k\} > x\}}{\overline{F}(x)} \\
=\;& (N-1)c_k + N\sum_{j\in A^{k,l}} c_j
\end{aligned}
$$

which together with (7) completes the proof. $\qquad\square$

Remark 4.3.2 Note that for station k with $|p(k)| = 1$, tail asymptotics of the waiting times only depends on the service time at station k and is the same as the tail asymptotics of waiting times in closed tandem queues (see Ayhan, Palmowski, and Schlegel [9]). However, for station k with $|p(k)| > 1$, the asymptotic tail distribution of the waiting time of the nth arriving customer from station $l \in p(k)$ at station k depends not only on the service time distribution of station k but also on the service time distributions of all the stations that belong to $A^{k,l}$.

4.3.2 Stationary Characteristics

Since the general fork and join network that we study is a (max,+) linear system (see Baccelli, Cohen, Olsder and Quadrat [11] for details of (max,+) linear systems), using the analysis in Section 7.5 of Baccelli, Cohen, Olsder and Quadrat [11], one can derive conditions under which the stationary characteristics exist. In particular, one can conclude from Theorem 7.94 of Baccelli, Cohen, Olsder and Quadrat [11] that if there exists a $k \in \{1, \ldots, K\}$ such that $N_{k,l} > 0$ for all $l \in p(k)$ and $B_k(\cdot)$ has infinite support (i.e., there exists a station which is ready to process at time 0 and has a service time distribution with infinite support), then the sequence of vectors $\{(X_n^k - X_{n-1}^j) : k, j \in \{1, \ldots, K\}\}_{n \geq 1}$ admits a unique stationary regime which is integrable, directly reachable, independent of the initial condition and $\{(X_n^k - X_{n-1}^j) : k, j \in \{1, \ldots, K\}\}$ couples with it in finite time. Thus, there exists a finite random variable T such that $X_n^k - X_{n-1}^k = Z_n$ for all $n \geq T$ (see the definition of coupling on page 87 of Baccelli and Brémaud [10]). Then for all $x \geq 0$,

$$\left| \mathbb{P}\{ \sum_{m=n+1}^{n+N} (X_m^k - X_{m-1}^k) \leq x \} - \mathbb{P}\{ \sum_{m=n+1}^{n+N} Z_m \leq x \} \right|$$

$$= \left| \mathbb{P}\{ \sum_{m=n+1}^{n+N} (X_m^k - X_{m-1}^k) \leq x, \ \sum_{m=n+1}^{n+N} (X_m^k - X_{m-1}^k) = \sum_{m=n+1}^{n+N} Z_m \} \right.$$

$$+ \mathbb{P}\{ \sum_{m=n+1}^{n+N} (X_m^k - X_{m-1}^k) \leq x, \ \sum_{m=n+1}^{n+N} (X_m^k - X_{m-1}^k) \neq \sum_{m=n+1}^{n+N} Z_m \}$$

$$- \mathbb{P}\{ \sum_{m=n+1}^{n+N} Z_m \leq x, \ \sum_{m=n+1}^{n+N} (X_m^k - X_{m-1}^k) = \sum_{m=n+1}^{n+N} Z_m \}$$

$$\left. - \mathbb{P}\{ \sum_{m=n+1}^{n+N} Z_m \leq x, \ \sum_{m=n+1}^{n+N} (X_m^k - X_{m-1}^k) \neq \sum_{m=n+1}^{n+N} Z_m \} \right|$$

$$= \left| \mathbb{P}\{ \sum_{m=n+1}^{n+N} (X_m^k - X_{m-1}^k) \leq x, \ \sum_{m=n+1}^{n+N} (X_m^k - X_{m-1}^k) \neq \sum_{m=n+1}^{n+N} Z_m \} \right.$$

$$\left. - \mathbb{P}\{ \sum_{m=n+1}^{n+N} Z_m \leq x, \ \sum_{m=n+1}^{n+N} (X_m^k - X_{m-1}^k) \neq \sum_{m=n+1}^{n+N} Z_m \} \right|$$

$$\leq \; 2\mathbb{P}\{ \sum_{m=n+1}^{n+N} (X_m^k - X_{m-1}^k) \neq \sum_{m=n+1}^{n+N} Z_m \} \leq 2P(T > n). \tag{8}$$

Since T is a finite random variable, it follows from (8) that

$$\lim_{n\to\infty} \left| \mathbb{P}\{C_n^k \leq x\} - \mathbb{P}\{C^k \leq x\} \right| = 0, \tag{9}$$

where C^k denotes the stationary cycle time at station $k \in \{1, \ldots, K\}$. Since the convergence in Proposition 4.3.1 is uniform, combining (9) with Proposition 4.3.1, we have the following result.

Proposition 4.3.3 *If there exists a $k \in \{1, \ldots, K\}$ such that $N_{k,l} > 0$ for all $l \in p(k)$ and $B_k(\cdot)$ has infinite support, then for all $k \in \{1, \ldots, K\}$*

$$\lim_{x\to\infty} \frac{\mathbb{P}\{C^k > x\}}{\overline{F}(x)} \; = \; N \sum_{j=1}^{K} c_j.$$

Let W^k denote the stationary waiting time at station k with $|p(k)| = 1$ and $W^{k,l}$ denote the stationary waiting time of an arbitrary customer at station k (with $|p(k)| > 1$) coming from station $l \in p(k)$. The following result can be obtained using a similar coupling argument since $\{(X_n^k - X_{n-1}^j) : k, j \in \{1, \ldots, K\}\}$ couples with a unique stationary regime in finite time under the assumption that there exists a $k \in \{1, \ldots, K\}$ such that $N_{k,l} > 0$ for all $l \in p(k)$ and $B_k(\cdot)$ has infinite support.

Proposition 4.3.4 *Suppose there exists a $k \in \{1, \ldots, K\}$ such that $N_{k,l} > 0$ for all $l \in p(k)$ and $B_k(\cdot)$ has infinite support. Then for $k \in \{1, \ldots, K\}$ with $|p(k)| = 1$ and $B_k \in \mathcal{S}$,*

$$\lim_{x\to\infty} \frac{\mathbb{P}\{W^k > x\}}{\overline{F}(x)} \; = \; (N-1)c_k$$

and for $k \in \{1, \ldots, K\}$ with $|p(k)| > 1$ and $l \in p(k)$, if $B_k \in \mathcal{S}$ or $B_j \in \mathcal{S}$ for some $j \in A^{k,l}$,

$$\lim_{x\to\infty} \frac{\mathbb{P}\{W^{k,l} > x\}}{\overline{F}(x)} \; = \; (N-1)c_k + N \sum_{j \in A^{k,l}} c_j.$$

Note that if the service times at all stations have infinite support (which is clearly satisfied if the service time distributions are subexponential), the above condition is satisfied and a stationary regime exists.

4.4   Numerical Results

In this section, we provide numerical experiments in order to understand how fast the convergence of tail probabilities of cycle times and waiting times is to their asymptotic counter parts.

We consider two systems which have the same structure of Figure 1 with $K = 7$ stations. However, for System 1, $N_k = 1$ for all $k \in \{1, \ldots, 7\}$ and thus, $N = 3$. For System 2, $N_k = 2$ for all $k \in \{1, \ldots, 7\}$ and thus, $N = 6$. Hence, Systems 1 and 2 have the same number of stations and the same structure but different number of customers in system. We assume that service time distribution at all stations for both systems is Pareto with parameter 1 (i.e., $\overline{B}_k(x) = x^{-1}$ for all $k \in \{1, \ldots, 7\}$). Then, from (2), constants c_k are equal to 1 for all $k \in \{1, \ldots, 7\}$. Since the tail asymptotics of transient cycle times and waiting times are independent of n as long as $n \geq 2N$ as discussed in Section 4.3, we consider the system time and waiting times of the 12th customer in both systems. Note that a system time is the cycle time corresponding to station 7. In particular, for each value of x, we first approximate the tail probabilities of system times and waiting times using the tail asymptotics of system times and waiting times as given in Proposition 4.3.1 and 4.3.2, respectively and compare them with the tail probabilities obtained from simulation analysis. In our simulation study, for each value of x we run 41 batches of 10,000 replications and compute the average and 95% confidence interval of the corresponding tail probability. However, for purposes of clarity, we do not present the confidence intervals in Figures 2 to 9.

First, we compute the tail probabilities of system times of the 12th customer for

both systems. Also, we compute the tail probabilities of waiting times at stations 1, 4, 6, and 7 for both systems. Note that stations 1 and 6 are the case with $|p(i)| = 1$ and stations 4 and 7 are the case with $|p(i)| > 1$. In particular, for waiting times at stations 4 and 7, we compute waiting time of the 12th arriving customer from stations 1 and 2 at station 4, and waiting time of the 12th arriving customer from stations 4, 5, and 6 at station 7. Figure 2 displays the tail probabilities of system times of the 12th customer for both systems and shows that the tail asymptotics could be used to approximate the tail probabilities of system times as x increases from medium to large values. In particular, when the total number of customers in the system is small (i.e., System 1), the convergence of the tail asymptotics to the actual tail probability is fast. Figures 3 to 9 present the tail probabilities of waiting times of the 12th customer at stations 1, 6, 4, and 7 for Systems 1 and 2. As the figures illustrate, the tail asymptotics could provide a good approximation for the tail probabilities of waiting times as x increases from medium to large values. Especially, the convergence of the tail asymptotics to the tail probabilities is fast for the waiting times at stations 1 and 6 (which are the cases with $|p(i)| = 1$). Note that the convergence of the tail asymptotics for waiting time of the 12th arriving customer from station 4 at station 7 is slower than the other two cases (coming from stations 5 and 6). This is because station 4 has more than one preceding station.
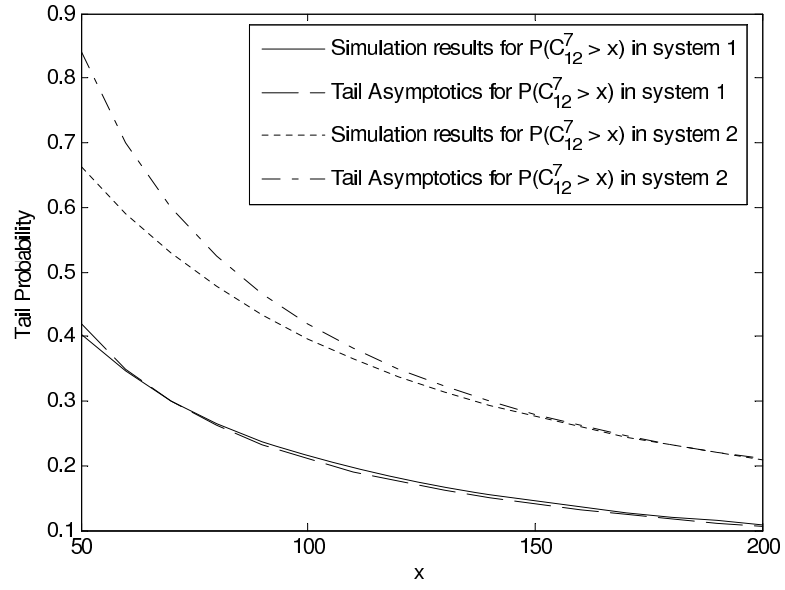
Figure 2: Simulation results and the tail asymptotics of system times for Systems 1 and 2
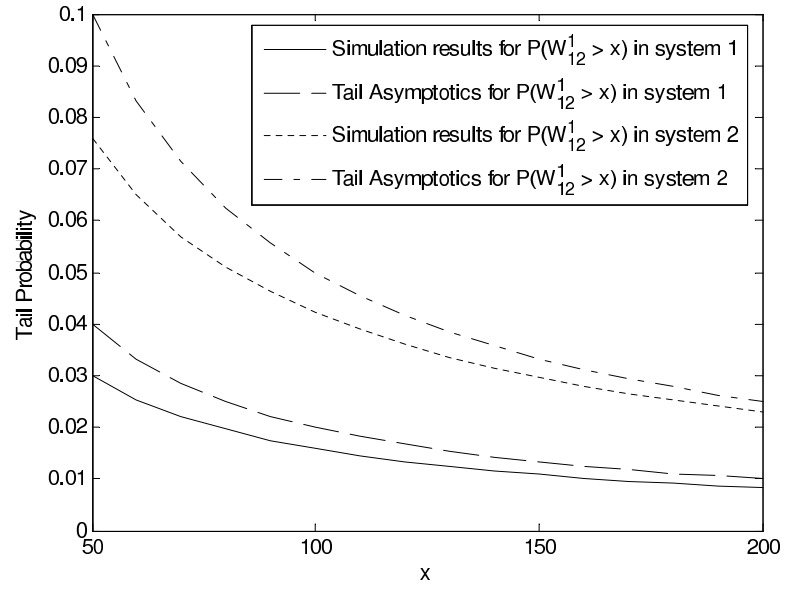
Figure 3: Simulation results and the tail asymptotics of waiting times at station 1 for Systems 1 and 2
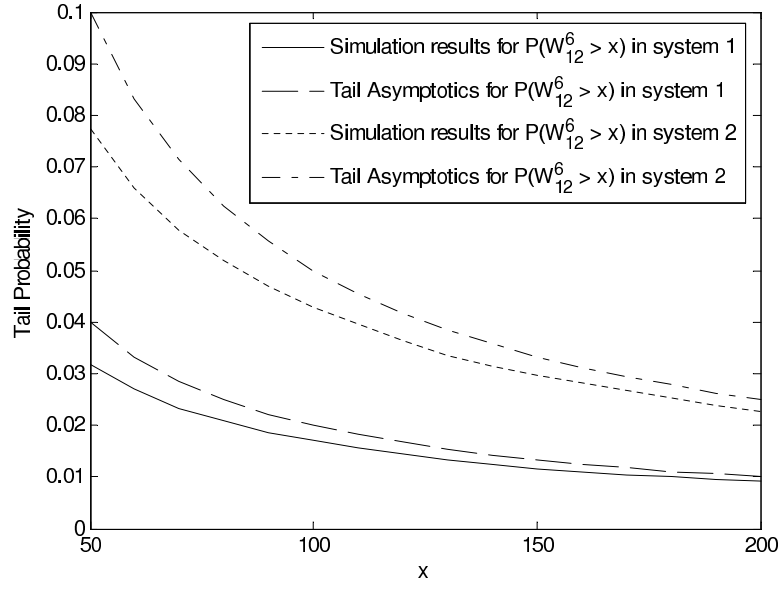
Figure 4: Simulation results and the tail asymptotics of waiting times at station 6 for Systems 1 and 2

Figure 5: Simulation results and the tail asymptotics of waiting times of the arriving customer from station 1 at station 4 for Systems 1 and 2

30

Figure 6: Simulation results and the tail asymptotics of waiting times of the arriving customer from station 2 at station 4 for Systems 1 and 2

Figure 7: Simulation results and the tail asymptotics of waiting times of the arriving customer from station 4 at station 7 for Systems 1 and 2

31
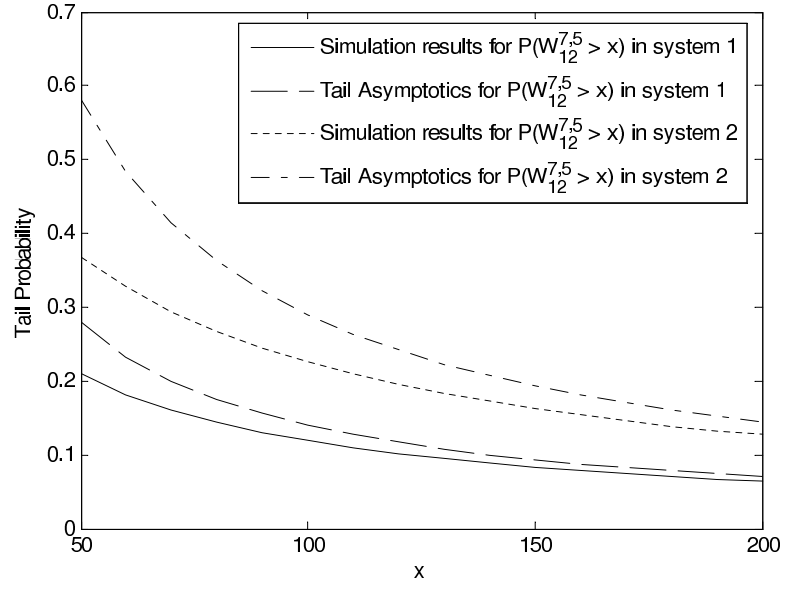
Figure 8: Simulation results and the tail asymptotics of waiting times of the arriving customer from station 5 at station 7 for Systems 1 and 2

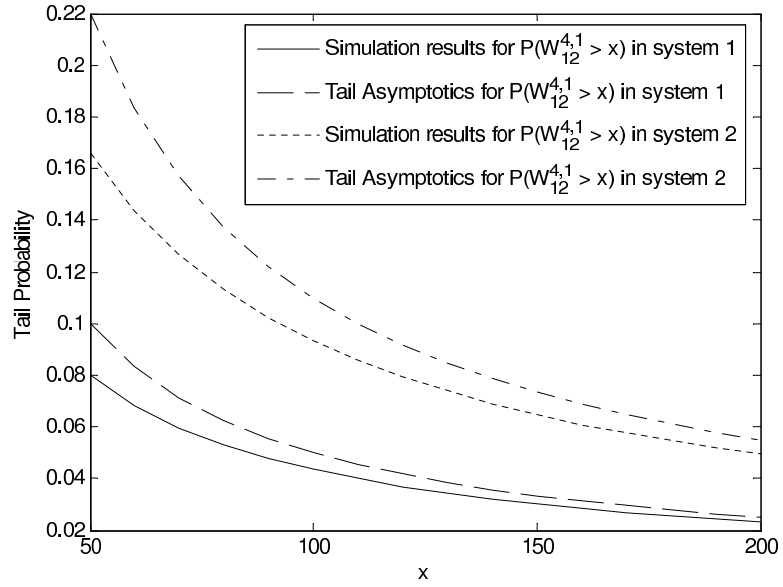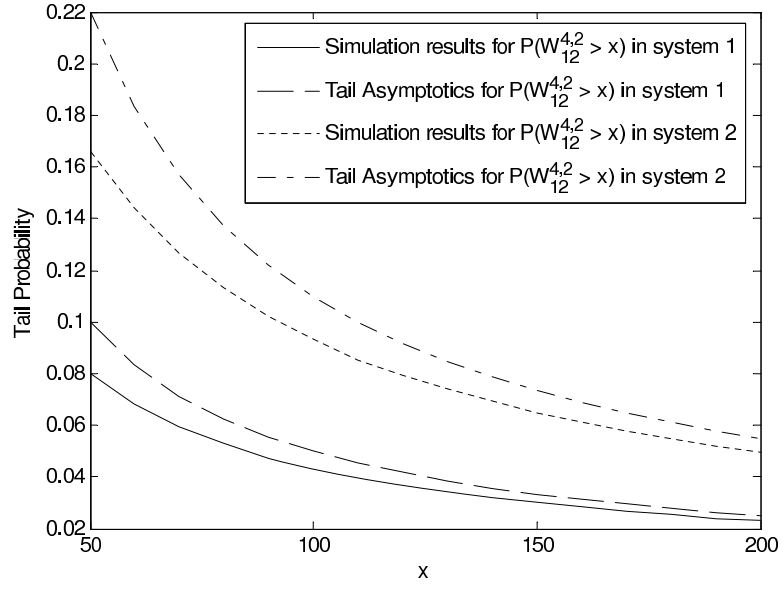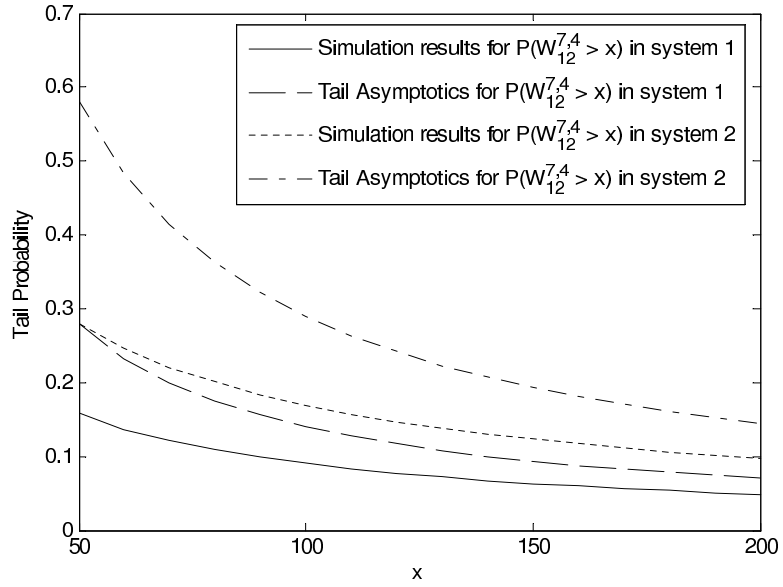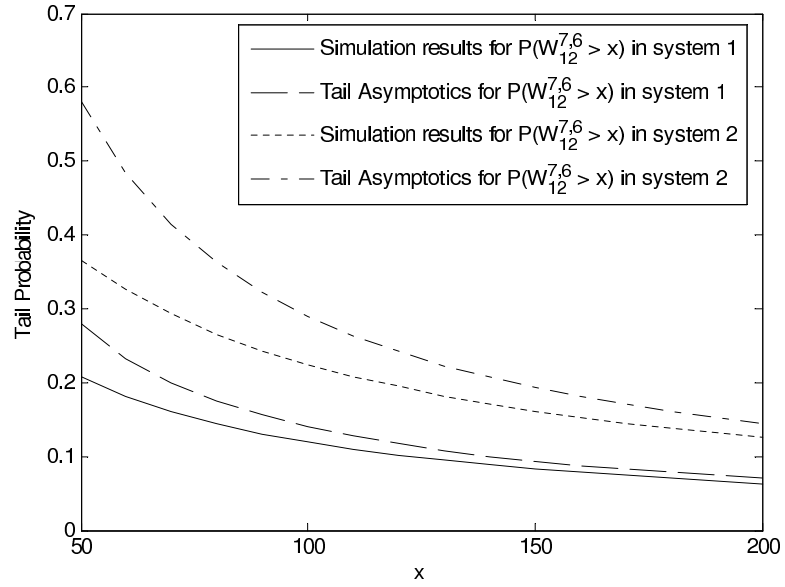**Figure 9:** Simulation results and the tail asymptotics of waiting times of the arriving customer from station 6 at station 7 for Systems 1 and 2

CHAPTER V

CYCLIC QUEUEING NETWORKS WITH SUBEXPONENTIAL SERVICE TIMES AND FINITE BUFFERS

5.1 Introduction

In this chapter, we study cyclic queueing networks with K stations ($K \geq 2$) and finite buffers as shown in Figure 10. There is a single server at each station $k \in \{1, \ldots, K\}$ and the service discipline at all stations is First Come First Served (FCFS). Since the buffer size of stations is finite, customers could be blocked. We analyze this system under communication blocking and manufacturing blocking schemes which are commonly encountered in practice; see for example Altiok and Stidham [1], Brandwajn and Jow [21] and Perros and Altiok [48]. In communication blocking, a server is not allowed to start service until space is available in the downstream buffer. On the other hand, in manufacturing blocking, at the time of service completion, the customer is not allowed to move to downstream buffer if that buffer is full. We are again interested in the tail asymptotics of transient and stationary cycle times and waiting times in this network.

Let M_k be the maximum allowable number of customers at station $k \in \{1, \ldots, K\}$ and $1 \leq M_k < \infty$. Thus, the total capacity of system is $\sum_{k=1}^{K} M_k$ which is denoted by M. Moreover, as defined in Chapter 4, let N_k be the initial number of customers at station $k \in \{1, \ldots, K\}$ and $N \geq 1$ be the total number of customers in the system. Then, $\sum_{k=1}^{K} N_k = N$. Clearly, $0 \leq N_k \leq M_k$ for all $k \in \{1, \ldots, K\}$. We assume that all stations are idle at time 0 and if there is a customer at a station, the service on that customer has not started before time 0. Note that N should be less than total

33

capacity M. Otherwise, nobody can move to the downstream station under both blocking schemes that we consider. Finally, $H_k = M_k - N_k$ denotes the number of empty spaces in the buffer of station $k \in \{1, \ldots, K\}$ at time 0. Clearly, $0 \leq H_k \leq M_k$ for all $k \in \{1, \ldots, K\}$. Service times at station $k \in \{1, \ldots, K\}$ are independent and identically distributed random variables $\{B_n^k\}$ with distribution function $B_k(\cdot)$. The sequence of service times at each station is independent of the service times at the other stations. Furthermore, we assume that there exists a subexponential distribution $F(\cdot)$ ($F \in S$) and there exist constants $c_k \in [0, \infty)$ with $\sum_{k=1}^{K} c_k > 0$ such that for all $k \in \{1, \ldots, K\}$,

$$\lim_{x \to \infty} \frac{\overline{B}_k(x)}{\overline{F}(x)} = c_k. \tag{10}$$
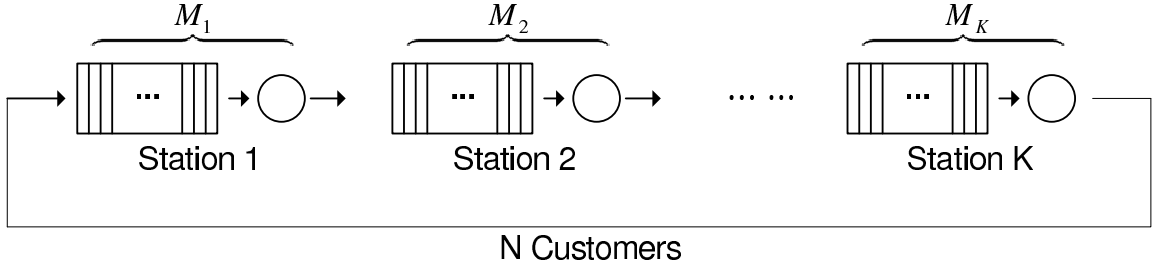
Figure 10: Cyclic queueing networks with K stations and finite buffers.

The remainder of this chapter is organized as follows. In Section 5.2, we focus on closed tandem queues with communication blocking. In particular, we provide preliminary results in Section 5.2.1 and investigate the tail asymptotics of the nth cycle time and waiting time at station k for all $k \in \{1, \ldots, K\}$ in Section 5.2.2. Similarly, in Section 5.3, we study the tail asymptotics of transient cycle times and waiting times in closed tandem queues with manufacturing blocking. Section 5.4 focuses on stationary cycle times and waiting times. Finally, we study the convergence behavior for the tail asymptotics of transient cycle times and waiting times in Section 5.5.

34

5.2   Communication Blocking

In this section, we consider cyclic queueing networks with communication blocking. This blocking requires a server not to initiate service of a customer if the downstream buffer is full. In this case, the server remains unavailable until the current service at the next server is completed.

5.2.1   Preliminaries

We introduce the following notation

$$[j] = \begin{cases} K & \text{if } j \bmod K = 0, \\ j \bmod K & \text{if } j \bmod K \neq 0. \end{cases}$$

For notational convenience, define

$$\mathcal{H}_{[k+1],[k+u]} = \sum_{i=1}^{u} H_{[k+i]},$$

with the convention that summation over a empty set is zero. Note that $\mathcal{H}_{[k+1],[k+u]}$ is the total number of available spaces from station $[k+1]$ to station $[k+u]$ (which is in the direction of customer flow) at time zero for all $u \geq 0$ and all $k \in \{1, \ldots, K\}$. Similarly, as in Chapter 4, define

$$\mathcal{N}_{[k],[k-u]} = \sum_{i=0}^{u} N_{[k-i]}$$

with the convention that summation over a empty set is zero. Note that $\mathcal{N}_{[k],[k-u]}$ is the total number of initial customers from station $[k]$ to station $[k-u]$ (which is in the opposite direction of customer flow for $u \geq 0$) and for all $k \in \{1, \ldots, K\}$. Unlike Chapter 4, there is only one path from station $[k]$ to $[k-u]$ due to tandem queues. Finally, for all $k \in \{1, \ldots, K\}$ and $u \geq 0$, define

$$\mathcal{V}_u^k = \min\{\mathcal{N}_{[k],[k-u+1]}, \mathcal{H}_{[k+1],[k+K-u]}\}.$$

As defined in Chapter 4, let X_n^k denote the departure time of the nth customer from station $k \in \{1, \ldots, K\}$. Then, we have the following expression.

Proposition 5.2.1 *For all $k \in \{1, \dots, K\}$ and $n \geq 1$,*

$$X_n^k = \max\{X_{n-1}^k + B_n^k, \ X_{n-N_k}^{[k-1]} + B_n^k, \ X_{n-H_{[k+1]}}^{[k+1]} + B_n^k\} \tag{11}$$

with the convention that $X_n^k = 0$ for all $n \leq 0$ and all $k \in \{1, \dots, K\}$.

Proof follows from the observation that the nth customer starts the service at station k at the time $\max\{X_{n-1}^k, \ X_{n-N_k}^{[k-1]}, \ X_{n-H_{[k+1]}}^{[k+1]}\}$ since a server at station k cannot initiate the service of the nth customer if the station $[k+1]$ is full. □

Remark 5.2.1

1. Let M_{k*} be the minimum of M_k for all $k \in \{1, \dots, K\}$. If $1 \leq N \leq M_{k*}$, there is no blocking because blocking occurs when the next station is full. Thus, (11) is reduced to $X_n^k = \max\{X_{n-1}^k + B_n^k, \ X_{n-N_k}^{[k-1]} + B_n^k\}$ for all $k \in \{1, \dots, K\}$, which matches the departure time expression of the infinite buffer case.

2. When $N \geq \sum_{u=1}^{K-1} M_{[k+u]} + 1$ for some station k, (11) is reduced to $X_n^k = \max\{X_{n-1}^k + B_n^k, \ X_{n-H_{[k+1]}}^{[k+1]} + B_n^k\}$ because $X_{n-N_k}^{[k-1]} \leq X_{n-1}^k$.

We develop upper and lower bounds on X_n^k for $k \in \{1, \dots, K\}$ in the following propositions.

Proposition 5.2.2 *For all $k \in \{1, \dots, K\}$ and $n \geq 1$,*

$$X_n^k \leq \sum_{u=0}^{K-1} \sum_{r=1}^{n-\mathcal{V}_u^k} B_r^{[k-u]}$$

with the convention that the summation over an empty set is zero.

Proof Clearly, the $(n - \mathcal{N}_{[k],[k-u+1]})^{th}$ customer needs to depart from station $[k-u]$ for $u = 0, \dots, K-1$ before X_n^k. At the time of joining the server of station $[k-u]$, if the downstream buffer is full, he or she needs to wait for at least one service completion

36

at station $[k-u+1]$. We can immediately obtain the following upper bound on X_n^k.

$$
\begin{aligned}
X_n^k &\leq \sum_{u=0}^{K-1} \sum_{r=1}^{n-\mathcal{N}_{[k],[k-u+1]}} \{B_r^{[k-u]} + \sum_{q=1}^{K-1} B_{r-\mathcal{H}_{[k-u+1],[k-u+q]}}^{[k-u+q]}\} \\
&= \sum_{u=0}^{K-1} \sum_{r=1}^{n-\mathcal{N}_{[k],[k-u+1]}} \sum_{q=0}^{K-1} B_{r-\mathcal{H}_{[k-u+1],[k-u+q]}}^{[k-u+q]} \\
&= \sum_{u=0}^{K-1} \sum_{q=0}^{K-1} \sum_{r=1}^{n-\mathcal{N}_{[k],[k-u+1]}} B_{r-\mathcal{H}_{[k-u+1],[k-u+q]}}^{[k-u+q]} \\
&= \sum_{u=0}^{K-1} \sum_{q=0}^{K-1} \sum_{r=1-\mathcal{H}_{[k-u+1],[k-u+q]}}^{n-\mathcal{N}_{[k],[k-u+1]}-\mathcal{H}_{[k-u+1],[k-u+q]}} B_r^{[k-u+q]} \\
&= \sum_{w=0}^{K-1} \sum_{r=1}^{n-\min\{\mathcal{N}_{[k],[k-w+1]},\mathcal{H}_{[k+1],[k+K-w]}\}} B_r^{[k-w]}.
\end{aligned}
\tag{12}
$$

The last equation following from setting $w = u - q$. $\qquad\square$

Proposition 5.2.3 *For all $k \in \{1,\ldots,K\}$ and $n \geq 1$,*

$$
X_n^k \geq \max_{u=0,\ldots,K-1} \{\sum_{r=1}^{n-\mathcal{V}_u^k} B_r^{[k-u]}\}
$$

with the convention that the summation over an empty set is zero and the maximization over an empty set is equal to $-\infty$.

Proof follows from the observation that before X_n^k at least $(n - \mathcal{V}_u^k)$ customers must have departed from station $[k-u]$ for all $u = 0,\ldots,K-1$. $\qquad\square$

5.2.2 Cycle Times and Waiting Times

As defined in Chapter 4, let C_n^k denote the nth cycle time at station $k \in \{1,\ldots,K\}$ which is the time between two successive departures of the same customer from station k. Thus, the nth cycle time at station $k \in \{1,\ldots,K\}$ is computed as

$$
C_n^k = X_{n+N}^k - X_n^k.
\tag{13}
$$

The next proposition provides the tail asymptotics of C_n^k for all $k \in \{1,\ldots,K\}$.

Proposition 5.2.4 *For all $k \in \{1, \ldots, K\}$ and all $n \geq \max_{u=0,\ldots,K-1} \mathcal{V}_u^k$,*

$$\lim_{x \to \infty} \frac{\mathbb{P}\{C_n^k > x\}}{\overline{F}(x)} = N \sum_{j=1}^{K} c_j$$

where the convergence is uniform in n.

Proof For the upper bound for C_n^k, define the set of service times $\mathcal{J}$ as

$$\mathcal{J} = \bigcup_{u=0}^{K-1} \bigcup_{r=n+1-\mathcal{V}_u^k}^{n+N-\mathcal{V}_u^k} B_r^{[k-u]}.$$

Note that at least one of the service times in $\mathcal{J}$ must be in progress at any time in the interval $[X_n^k, X_{n+N}^k]$ and there is no other service time (other than those in $\mathcal{J}$) that could take place in this time interval and could have an effect on C_n^k. Then, we have

$$C_n^k \leq \sum_{u=0}^{K-1} \sum_{r=n+1-\mathcal{V}_u^k}^{n+N-\mathcal{V}_u^k} B_r^{[k-u]}. \tag{14}$$

Hence, from Corollary 3.0.1, for all $n \geq \max_{u=0,\ldots,K-1} \mathcal{V}_u^k$

$$
\begin{aligned}
\limsup_{x \to \infty} \frac{\mathbb{P}\{C_n^k > x\}}{\overline{F}(x)} &\leq \limsup_{x \to \infty} \frac{\mathbb{P}\{\sum_{u=0}^{K-1} \sum_{r=1}^{N} B_r^{[k-u]} > x\}}{\overline{F}(x)} \\
&= N \sum_{j=1}^{K} c_j. \tag{15}
\end{aligned}
$$

Now we provide a lower bound on C_n^k. Note that all service times in $\mathcal{J}$ take place within the time interval $[X_{n-N}^k, X_{n+N}^k]$. Let

$$
\begin{aligned}
T_n^k &= X_n^k - X_{n-N}^k \\
&\leq \sum_{u=0}^{K-1} \sum_{r=n-N+1-\mathcal{V}_u^k}^{n-\mathcal{V}_u^k} B_r^{[k-u]}
\end{aligned}
$$

and observe that the service times that occur in the interval from X_{n-N}^k to X_n^k do not have an effect on C_n^k. Therefore,

$$
\begin{aligned}
C_n^k &\geq \max_{u=0,\ldots,K-1} \{ \max_{r=n+1-\mathcal{V}_u^k,\ldots,n+N-\mathcal{V}_u^k} B_r^{[k-u]} \} - T_n^k \\
&\geq \max_{u=0,\ldots,K-1} \{ \max_{r=n+1-\mathcal{V}_u^k,\ldots,n+N-\mathcal{V}_u^k} B_r^{[k-u]} \} - \sum_{u=0}^{K-1} \sum_{r=n-N+1-\mathcal{V}_u^k}^{n-\mathcal{V}_u^k} B_r^{[k-u]}.
\end{aligned}
$$

Thus, from Lemmas 3.0.3 and 3.0.4 and Corollary 3.0.1, for $n \geq \max_{u=0,\ldots,K-1} \mathcal{V}_u^k$, we have

$$
\begin{aligned}
& \liminf_{x \to \infty} \frac{\mathbb{P}\{C_n^k > x\}}{\overline{F}(x)} \\
\geq \ & \liminf_{x \to \infty} \frac{\mathbb{P}\{\max_{u=0,\ldots,K-1}\{\max_{r=N+1,\ldots,2N} B_r^{[k-u]}\} - \sum_{u=0}^{K-1} \sum_{r=1}^{N} B_r^{[k-u]} > x\}}{\overline{F}(x)} \\
= \ & N \sum_{j=1}^{K} c_j
\end{aligned}
$$

which together with (15) completes the proof. $\qquad \square$

Remark 5.2.2 The tail asymptotics of cycle times is the same as one in the tandem queue with infinite buffers as given in Ayhan, Palmowski and Schlegel [9].

As defined in Chapter 4, let W_n^k denote the waiting time of the nth customer of station $k \in \{1, \ldots, K\}$. Thus, W_n^k is the time from the arrival of the nth customer at station k until joining the server at that station and it is computed as

$$
W_n^k = \max\{X_{n-1}^k - X_{n-N_k}^{[k-1]}, 0\}. \tag{16}
$$

Define $\beta_q^k = \max\{(N-1) - \sum_{j=1}^{q} M_{[k+j]}, 0\}$ and $u^* = \min\{u : N \leq \sum_{j=0}^{u} M_{[k+j]}, u = 1, \ldots, K-1\}$. Consider the following assumption.

A.1. $B_j \in S$ for some $j \in \{k, \ldots, [k+u^*-1]\}$ or for some $j \in \{[k+u^*], \ldots, [k+q]\}$ where q is such that $\beta_q^k > 0$ and $\beta_{q+1}^k = 0$ for $q = u^*, \ldots, K-1$.

The next proposition gives the tail asymptotics of the nth waiting time at station $k \in \{1, \ldots, K\}$.

Proposition 5.2.5 *For all* $k \in \{1, \ldots, K\}$ *with assumption A.1. and* $n \geq \min\{N, M_k\}$,

$$
\lim_{x \to \infty} \frac{\mathbb{P}\{W_n^k > x\}}{\overline{F}(x)} = (\min\{N, M_k\} - 1) \sum_{j=0}^{u^*-1} c_{[k+j]} + \sum_{q=u^*}^{K-1} \beta_q^k c_{[k+q]},
$$

where the convergence is uniform in n.

Proof Note that W_n^k will attain its largest value if there are $(\min\{N, M_k\} - 1)$ customers waiting their service at station k at the time that the nth customer joins station k. That is the $(n - \min\{N, M_k\} + 1)^{th}$ customer is still in service at station k. Moreover, all $(\min\{N, M_k\} - 1)$ customers get blocked when the server initiate their service at station k. This argument immediately gives the following set of service times.

$$\mathcal{K} = \bigcup_{r=n-\min\{N,M_k\}+1}^{n-1} B_r^k \cup \bigcup_{u=1}^{u^*-1} \bigcup_{r=n-M_k+1-\mathcal{H}_{[k+1],[k+u]}}^{n-1-\mathcal{H}_{[k+1],[k+u]}} B_r^{[k+u]} \cup \bigcup_{q=u^*}^{K-1} \bigcup_{r=n-\beta_q^k-\mathcal{H}_{[k+1],[k+q]}}^{n-1-\mathcal{H}_{[k+1],[k+q]}} B_r^{[k+q]},$$

where $u^* = \min\{u : N \leq \sum_{j=0}^{u} M_{[k+j]}, u = 1, \ldots, K - 1\}$.

Note that at least one of the service times in $\mathcal{K}$ must be in progress on station k at anytime in the time interval $[X_{n-N_k}^{[k-1]}, X_{n-1}^k]$ if $X_{n-1}^k \geq X_{n-N_k}^{[k-1]}$ (otherwise $W_n^k = 0$) and there is no other service time (other than those in $\mathcal{K}$) that could take place on station k in this time interval and could have an effect on W_n^k. Then,

$$W_n^k \leq \sum_{r=n-\min\{N,M_k\}+1}^{n-1} B_r^k + \sum_{u=1}^{u^*-1} \sum_{r=n-M_k+1-\mathcal{H}_{[k+1],[k+u]}}^{n-1-\mathcal{H}_{[k+1],[k+u]}} B_r^{[k+u]} + \sum_{q=u^*}^{K-1} \sum_{r=n-\beta_q^k-\mathcal{H}_{[k+1],[k+q]}}^{n-1-\mathcal{H}_{[k+1],[k+q]}} B_r^{[k+q]}.$$

Hence, from Corollary 3.0.1, for $n \geq \min\{N, M_k\}$

$$\limsup_{x\to\infty} \frac{\mathbb{P}\{W_n^k > x\}}{\overline{F}(x)}$$

$$\leq \limsup_{x\to\infty} \frac{\mathbb{P}\{\sum_{r=1}^{\min\{N,M_k\}-1} B_r^k + \sum_{u=1}^{u^*-1}\sum_{r=1}^{M_k-1} B_r^{[k+u]} + \sum_{q=u^*}^{K-1}\sum_{r=1}^{\beta_q^k} B_r^{[k+q]} > x\}}{\overline{F}(x)}$$

$$= (\min\{N, M_k\} - 1)c_k + (M_k - 1)\sum_{u=1}^{u^*-1} c_{[k+u]} + \sum_{q=u^*}^{K-1} \beta_q^k c_{[k+q]}. \qquad (17)$$

Next we obtain a lower bound on W_n^k. Note that all service times in $\mathcal{K}$ take place at station k in the time interval $[X_{n-N}^k, X_{n-1}^k]$. Moreover, completed service times that take place in the interval from X_{n-N}^k to $X_{n-N_k}^{[k-1]}$ do not have an effect on W_n^k. We have $X_{n-N}^k \geq X_{n-N-N_k}^{[k-1]}$ from the recursive expression of the departure times in (11). As

in the proof of Proposition 5.2.4, let

$$
\begin{aligned}
T^{[k-1]}_{n-N_k} &= X^{[k-1]}_{n-N_k} - X^{[k-1]}_{n-N-N_k} \\
&\leq \sum_{u=0}^{K-1} \sum_{r=n-N-N_k+1-\mathcal{V}^{[k-1]}_u}^{n-N_k-\mathcal{V}^{[k-1]}_u} B^{[k-1-u]}_r \\
&\leq \sum_{u=0}^{K-2} \sum_{r=n-N-N_k+1-\mathcal{V}^{[k-1]}_u}^{n-N_k-\mathcal{V}^{[k-1]}_u} B^{[k-1-u]}_r + \sum_{r=n-N-N_k+1-\mathcal{V}^{[k-1]}_{K-1}}^{n-N_k-\mathcal{V}^{[k-1]}_{K-1}} B^k_r.
\end{aligned}
$$

Since the upper bound on $T^{[k-1]}_{n-N_k}$ is independent of the service times in $\mathcal{K}$, we have

$$
\begin{aligned}
W^k_n \geq \max\Bigg\{ &\max_{r=n-\min\{N,M_k\}+1,\dots,n-1} B^k_r + \max_{u=1,\dots,u^*-1} \max_{r=n-M_k+1-\mathcal{H}_{[k+1],[k+u]},\dots,n-1-\mathcal{H}_{[k+1],[k+u]}} B^{[k+u]}_r \\
&+ \max_{q=u^*,\dots,K-1} \max_{r=n-\beta^k_q-\mathcal{H}_{[k+1],[k+q]},\dots,n-1-\mathcal{H}_{[k+1],[k+q]}} B^{[k+q]}_r \Bigg\} \\
&- \sum_{u=0}^{K-2} \sum_{r=n-N-N_k+1-\mathcal{V}^{[k-1]}_u}^{n-N_k-\mathcal{V}^{[k-1]}_u} B^{[k-1-u]}_r - \sum_{r=n-N-N_k+1-\mathcal{V}^{[k-1]}_{K-1}}^{n-N_k-\mathcal{V}^{[k-1]}_{K-1}} B^k_r
\end{aligned}
$$

and then from Lemma 3.0.3 and 3.0.4, for $n \geq \min\{N, M_k\}$, we have

$$
\begin{aligned}
&\liminf_{x\to\infty} \frac{\mathbb{P}\{W^k_n > x\}}{\overline{F}(x)} \\
&\geq \liminf_{x\to\infty} \mathbb{P}\Bigg\{ \max\Big\{ \max_{r=1,\dots,\min\{N,M_k\}-1} B^k_r, \max_{u=1,\dots,u^*-1} \max_{r=1,\dots,M_k-1} B^{[k+u]}_r, \\
&\qquad \max_{q=u^*,\dots,K-1} \max_{r=1,\dots,\beta^k_q} B^{[k+u]}_r \Big\} > x \Bigg\} \Big/ \overline{F}(x) \\
&= (\min\{N, M_k\} - 1)c_k + (M_k - 1)\sum_{u=1}^{u^*-1} c_{[k+u]} + \sum_{q=u^*}^{K-1} \beta^k_q c_{[k+q]}
\end{aligned}
$$

which together with (17) completes the proof. □

5.3   Manufacturing Blocking

We now consider the system described in Section 5.1 while it is operating under the manufacturing blocking rule. Under this blocking scheme, at the completion of a service at station k, the customer moves into the buffer of station $k+1$, if that buffer is not full. Otherwise it has to wait with the server at station k until the downstream

buffer has a free space.

We first provide the recursive expression for the departure time of the nth customer from station k for all $k \in \{1, \ldots, K\}$.

Proposition 5.3.1 *For all $k \in \{1, \ldots, K\}$ and $n \geq 1$,*

$$X_n^k = \max\{X_{n-1}^k + B_n^k, \ X_{n-N_k}^{[k-1]} + B_n^k, \ X_{n-H_{[k+1]}}^{[k+1]}\} \tag{18}$$

where $X_n^k = 0$ for all $n \leq 0$ and all $k \in \{1, \ldots, K\}$.

Proof At the time of the nth service completion at station $k \in \{1, \ldots, K\}$, if buffer $[k+1]$ is not full, the nth departure time is the maximum of $X_{n-1}^k + B_n^k$ and $X_{n-N_k}^{[k-1]} + B_n^k$. If buffer $[k+1]$ is full, the nth customer is blocked and needs to wait for the blocking to be cleared. □

By summing up all the service times that appear in (18), one can easily see the following upper bound on X_n^k.

Proposition 5.3.2 *For all $k \in \{1, \ldots, K\}$ and $n \geq 1$,*

$$X_n^k \leq \sum_{u=0}^{K-1} \sum_{r=1}^{n-\mathcal{V}_u^k} B_r^{[k-u]}$$

with the convention that the summation over an empty set is zero.

One can also obtain the following lower bound on X_n^k from the observation that before X_n^k at least $(n - \mathcal{V}_u^k)$ customers must have departed from station $[k - u]$ for $u = 0, \ldots, K - 1$.

Proposition 5.3.3 *For all $k \in \{1, \ldots, K\}$ and $n \geq 1$,*

$$X_n^k \geq \max_{u=0,\ldots,K-1}\{\sum_{r=1}^{n-\mathcal{V}_u^k} B_r^{[k-u]}\}$$

with the convention that the summation over an empty set is zero and the maximization over an empty set is equal to $-\infty$.

Now, we provide the tail asymptotics of transient cycle times and waiting times on closed tandem queues with manufacturing blocking.

Proposition 5.3.4 *For all $k \in \{1, \ldots, K\}$ and all $n \geq \max_{u=0,\ldots,K-1} \mathcal{V}_u^k$,*

$$\lim_{x \to \infty} \frac{\mathbb{P}\{C_n^k > x\}}{\overline{F}(x)} = N \sum_{j=1}^{K} c_j$$

where the convergence is uniform in n.

Proof We first provide the upper bound for C_n^k. Define the set of service times $\mathcal{J}'$ as

$$\mathcal{J}' = \bigcup_{u=0}^{K-1} \bigcup_{r=n+1-\mathcal{V}_u^k}^{n+N-\mathcal{V}_u^k} B_r^{[k-u]}.$$

Note that at least one of the service times in $\mathcal{J}'$ must be in progress at any time in the interval $[X_n^k, X_{n+N}^k]$ and there is no other service time (other than those in $\mathcal{J}'$) that could take place in this time interval and could have an effect on C_n^k. Then, we have

$$C_n^k \leq \sum_{u=0}^{K-1} \sum_{r=n+1-\mathcal{V}_u^k}^{n+N-\mathcal{V}_u^k} B_r^{[k-u]}. \tag{19}$$

Hence, from Corollary 3.0.1, for all $n \geq \max_{u=0,\ldots,K-1} \mathcal{V}_u^k$,

$$\limsup_{x \to \infty} \frac{\mathbb{P}\{C_n^k > x\}}{\overline{F}(x)} \leq \limsup_{x \to \infty} \frac{\mathbb{P}\{\sum_{u=0}^{K-1} \sum_{r=1}^{N} B_r^{[k-u]} > x\}}{\overline{F}(x)}$$

$$= N \sum_{j=1}^{K} c_j. \tag{20}$$

Next, we provide a lower bound on C_n^k. Note that all service times in $\mathcal{J}'$ take place within the time interval $[X_{n-N}^k, X_{n+N}^k]$. Let

$$T_n^k = X_n^k - X_{n-N}^k$$

$$\leq \sum_{u=0}^{K-1} \sum_{r=n-N+1-\mathcal{V}_u^k}^{n-\mathcal{V}_u^k} B_r^{[k-u]}$$

43

where the last inequality follows from the (13) and (19), and observe that the service times that occur in the interval from X_{n-N}^k to X_n^k do not have an effect on C_n^k. Therefore,

$$
\begin{aligned}
C_n^k &\geq \max_{u=0,\dots,K-1}\left\{\max_{r=n+1-\mathcal{V}_u^k,\dots,n+N-\mathcal{V}_u^k} B_r^{[k-u]}\right\} - T_n^k \\
&\geq \max_{u=0,\dots,K-1}\left\{\max_{r=n+1-\mathcal{V}_u^k,\dots,n+N-\mathcal{V}_u^k} B_r^{[k-u]}\right\} - \sum_{u=0}^{K-1}\sum_{r=n-N+1-\mathcal{V}_u^k}^{n-\mathcal{V}_u^k} B_r^{[k-u]}.
\end{aligned}
$$

Thus, from Lemmas 3.0.3 and 3.0.4 and Corollary 3.0.1, for $n \geq \max_{u=0,\dots,K-1} \mathcal{V}_u^k$, we have

$$
\begin{aligned}
&\liminf_{x\to\infty} \frac{\mathbb{P}\{C_n^k > x\}}{\overline{F}(x)} \\
&\geq \liminf_{x\to\infty} \frac{\mathbb{P}\{\max_{u=0,\dots,K-1}\{\max_{r=N+1,\dots,2N} B_r^{[k-u]}\} - \sum_{u=0}^{K-1}\sum_{r=1}^{N} B_r^{[k-u]} > x\}}{\overline{F}(x)} \\
&= N\sum_{j=1}^{K} c_j
\end{aligned}
$$

which together with (20) completes the proof. $\qquad\square$

Proposition 5.3.5 *For all $k \in \{1,\dots,K\}$ with assumption A.1. and $n \geq \min\{N, M_k\}$,*

$$
\lim_{x\to\infty} \frac{\mathbb{P}\{W_n^k > x\}}{\overline{F}(x)} = (\min\{N, M_k\} - 1)\sum_{j=0}^{u^*-1} c_{[k+j]} + \sum_{q=u^*}^{K-1} \beta_q^k c_{[k+q]},
$$

where the convergence is uniform in n.

Proof Note that W_n^k will attain its largest value if there are $(\min\{N, M_k\} - 1)$ customers waiting their service at station k at the time that the nth customer joins station k. That is the $(n - \min\{N, M_k\} + 1)^{th}$ customer is still in service at station k. In addition, all $(\min\{N, M_k\} - 1)$ customers get blocked when they depart from station k. This argument immediately gives the following set of service times.

$$
\mathcal{K}' = \bigcup_{r=n-\min\{N,M_k\}+1}^{n-1} B_r^k \cup \bigcup_{u=1}^{u^*-1} \bigcup_{r=n-M_k+1-\mathcal{H}_{[k+1],[k+u]}}^{n-1-\mathcal{H}_{[k+1],[k+u]}} B_r^{[k+u]} \cup \bigcup_{q=u^*}^{K-1} \bigcup_{r=n-\beta_q^k-\mathcal{H}_{[k+1],[k+q]}}^{n-1-\mathcal{H}_{[k+1],[k+q]}} B_r^{[k+q]},
$$

where $u^* = \min\{u : N \leq \sum_{j=0}^{u} M_{[k+j]}, u = 1, \ldots, K-1\}$.

Note that at least one of the service times in $\mathcal{K}'$ must be in progress on station k at anytime in the time interval $[X_{n-N_k}^{[k-1]}, X_{n-1}^k]$ if $X_{n-1}^k \geq X_{n-N_k}^{[k-1]}$ (otherwise $W_n^k = 0$) and there is no other service time (other than those in $\mathcal{K}'$) that could take place on station k in this time interval and could have an effect on W_n^k. Then,

$$W_n^k \leq \sum_{r=n-\min\{N,M_k\}+1}^{n-1} B_r^k + \sum_{u=1}^{u^*-1} \sum_{r=n-M_k+1-\mathcal{H}_{[k+1],[k+u]}}^{n-1-\mathcal{H}_{[k+1],[k+u]}} B_r^{[k+u]} + \sum_{q=u^*}^{K-1} \sum_{r=n-\beta_q^k-\mathcal{H}_{[k+1],[k+q]}}^{n-1-\mathcal{H}_{[k+1],[k+q]}} B_r^{[k+q]}.$$

Thus, from Corollary 3.0.1, for $n \geq \min\{N, M_k\}$

$$\limsup_{x \to \infty} \frac{\mathbb{P}\{W_n^k > x\}}{\overline{F}(x)}$$

$$\leq \limsup_{x \to \infty} \frac{\mathbb{P}\{\sum_{r=1}^{\min\{N,M_k\}-1} B_r^k + \sum_{u=1}^{u^*-1} \sum_{r=1}^{M_k-1} B_r^{[k+u]} + \sum_{q=u^*}^{K-1} \sum_{r=1}^{\beta_q^k} B_r^{[k+q]} > x\}}{\overline{F}(x)}$$

$$= (\min\{N, M_k\} - 1)c_k + (M_k - 1)\sum_{u=1}^{u^*-1} c_{[k+u]} + \sum_{q=u^*}^{K-1} \beta_q^k c_{[k+q]}. \tag{21}$$

Next we obtain a lower bound on W_n^k. Note that all service times in $\mathcal{K}'$ take place at station k in the time interval $[X_{n-N}^k, X_{n-1}^k]$. In addition, completed service times that take place in the interval from X_{n-N}^k to $X_{n-N_k}^{[k-1]}$ do not have an effect on W_n^k. We have $X_{n-N}^k \geq X_{n-N-N_k}^{[k-1]}$ from the recursive expression of the departure times in (18). As in the proof of Proposition 5.3.4, let

$$T_{n-N_k}^{[k-1]} = X_{n-N_k}^{[k-1]} - X_{n-N-N_k}^{[k-1]}$$

$$\leq \sum_{u=0}^{K-1} \sum_{r=n-N-N_k+1-\mathcal{V}_u^{[k-1]}}^{n-N_k-\mathcal{V}_u^{[k-1]}} B_r^{[k-1-u]}$$

$$\leq \sum_{u=0}^{K-2} \sum_{r=n-N-N_k+1-\mathcal{V}_u^{[k-1]}}^{n-N_k-\mathcal{V}_u^{[k-1]}} B_r^{[k-1-u]} + \sum_{r=n-N-N_k+1-\mathcal{V}_{K-1}^{[k-1]}}^{n-N_k-\mathcal{V}_{K-1}^{[k-1]}} B_r^k.$$

Since the upper bound on $T_{n-N_k}^{[k-1]}$ is independent of the service times in $\mathcal{K}'$, we have

$$W_n^k \geq \max\left\{ \max_{r=n-\min\{N,M_k\}+1,\ldots,n-1} B_r^k + \max_{u=1,\ldots,u^*-1} \max_{r=n-M_k+1-\mathcal{H}_{[k+1],[k+u]},\ldots,n-1-\mathcal{H}_{[k+1],[k+u]}} B_r^{[k+u]} \right.$$

$$\left. + \max_{q=u^*,\ldots,K-1} \max_{r=n-\beta_q^k-\mathcal{H}_{[k+1],[k+q]},\ldots,n-1-\mathcal{H}_{[k+1],[k+q]}} B_r^{[k+q]} \right\}$$

$$- \sum_{u=0}^{K-2} \sum_{r=n-N-N_k+1-\mathcal{V}_u^{[k-1]}}^{n-N_k-\mathcal{V}_u^{[k-1]}} B_r^{[k-1-u]} - \sum_{r=n-N-N_k+1-\mathcal{V}_{K-1}^{[k-1]}}^{n-N_k-\mathcal{V}_{K-1}^{[k-1]}} B_r^k$$

and then from Lemma 3.0.3 and 3.0.4, for $n \geq \min\{N, M_k\}$, we have

$$\liminf_{x\to\infty} \frac{\mathbb{P}\{W_n^k > x\}}{\overline{F}(x)}$$

$$\geq \liminf_{x\to\infty} \mathbb{P}\left\{ \max\{ \max_{r=1,\ldots,\min\{N,M_k\}-1} B_r^k, \max_{u=1,\ldots,u^*-1} \max_{r=1,\ldots,M_k-1} B_r^{[k+u]}, \right.$$

$$\left. \max_{q=u^*,\ldots,K-1} \max_{r=1,\ldots,\beta_q^k} B_r^{[k+u]} \} > x \right\} \Big/ \overline{F}(x)$$

$$= (\min\{N, M_k\} - 1)c_k + (M_k - 1)\sum_{u=1}^{u^*-1} c_{[k+u]} + \sum_{q=u^*}^{K-1} \beta_q^k c_{[k+q]}$$

which together with (21) completes the proof. $\square$

From Propositions 5.3.4 and 5.3.5, one can see that the tail asymptotics of cycle times and waiting times do not depend on the blocking schemes.

5.4 Stationary Cycle times and Waiting times

In this section, we provide the tail asymptotics of stationary cycle times and waiting times under both blocking schemes. Let C^k and W^k denote the stationary cycle time and waiting time at station $k \in \{1, \ldots, K\}$, respectively when they exist. Since the convergence in Propositions 5.2.4, 5.2.5, 5.3.4, and 5.3.5 is uniform in n, we immediately have the following result.

Proposition 5.4.1 *If a stationary regime exists, then under communication blocking and manufacturing blocking, for all $k \in \{1, \ldots, K\}$,*

$$\lim_{x\to\infty} \frac{\mathbb{P}\{C^k > x\}}{\overline{F}(x)} = N \sum_{j=1}^{K} c_j,$$

and for all $k \in \{1, \ldots, K\}$ with assumption A.1.,

$$\lim_{x \to \infty} \frac{\mathbb{P}\{W^k > x\}}{\overline{F}(x)} = (\min\{N, M_k\} - 1) \sum_{j=0}^{u^*-1} c_{[k+j]} + \sum_{q=u^*}^{K-1} \beta_q^k c_{[k+q]}.$$

We next provide sufficient conditions under which a stationary regime exists. Since a tandem queue with finite buffers is an example of a (max,+) linear system, as in Chapter 4, we can use the analysis in Section 7.5 of Baccelli, Cohen, Olsder and Quadrat [11]. From Theorem 7.94 of Baccelli, Cohen, Olsder and Quadrat [11], one can conclude that if there exists a station $k \in \{1, \ldots, K\}$ which is ready to start to process at time zero and has a service time distribution with infinite support, then the sequence of vectors $\{(X_n^k - X_{n-1}^j) : k, j \in \{1, \ldots, K\}\}_{n \geq 1}$ admits a unique stationary regime which is integrable, directly reachable, independent of the initial condition and couple with it in finite time. Hance, there exists a finite random variable T such that $X_n^k - X_{n-1}^k = Z_n$ for all $n \geq T$ (see Baccelli and Brémaud [10]). Then for all $x \geq 0$,

$$\left| \mathbb{P}\{ \sum_{m=n+1}^{n+N} (X_m^k - X_{m-1}^k) \leq x \} - \mathbb{P}\{ \sum_{m=n+1}^{n+N} Z_m \leq x \} \right|$$

$$= \left| \mathbb{P}\{ \sum_{m=n+1}^{n+N} (X_m^k - X_{m-1}^k) \leq x, \sum_{m=n+1}^{n+N} (X_m^k - X_{m-1}^k) = \sum_{m=n+1}^{n+N} Z_m \} \right.$$

$$+ \mathbb{P}\{ \sum_{m=n+1}^{n+N} (X_m^k - X_{m-1}^k) \leq x, \sum_{m=n+1}^{n+N} (X_m^k - X_{m-1}^k) \neq \sum_{m=n+1}^{n+N} Z_m \}$$

$$- \mathbb{P}\{ \sum_{m=n+1}^{n+N} Z_m \leq x, \sum_{m=n+1}^{n+N} (X_m^k - X_{m-1}^k) = \sum_{m=n+1}^{n+N} Z_m \}$$

$$\left. - \mathbb{P}\{ \sum_{m=n+1}^{n+N} Z_m \leq x, \sum_{m=n+1}^{n+N} (X_m^k - X_{m-1}^k) \neq \sum_{m=n+1}^{n+N} Z_m \} \right|$$

$$= \left| \mathbb{P}\{ \sum_{m=n+1}^{n+N} (X_m^k - X_{m-1}^k) \leq x, \sum_{m=n+1}^{n+N} (X_m^k - X_{m-1}^k) \neq \sum_{m=n+1}^{n+N} Z_m \} \right.$$

$$\left. - \mathbb{P}\{ \sum_{m=n+1}^{n+N} Z_m \leq x, \sum_{m=n+1}^{n+N} (X_m^k - X_{m-1}^k) \neq \sum_{m=n+1}^{n+N} Z_m \} \right|$$

$$\leq 2\mathbb{P}\{ \sum_{m=n+1}^{n+N} (X_m^k - X_{m-1}^k) \neq \sum_{m=n+1}^{n+N} Z_m \} \leq 2P(T > n). \tag{22}$$

47

Since T is a finite random variable, it follows from (22) that

$$\lim_{n\to\infty} \left| \mathbb{P}\{C_n^k \leq x\} - \mathbb{P}\{C^k \leq x\} \right| = 0. \tag{23}$$

The stationary waiting time can be obtained using a similar coupling argument.

Then, the sufficient condition for communication blocking is: there exists a $k \in \{1, \ldots, K\}$ such that $N_k > 0$, $H_{[k+1]} \geq 1$ and $B_k(\cdot)$ has infinite support.

Similarly, the sufficient condition for manufacturing blocking is: there exists a $k \in \{1, \ldots, K\}$ such that $N_k > 0$ and $B_k(\cdot)$ has infinite support.

Note that if the service times at all stations have infinite support, above sufficient conditions are satisfied and a stationary regime exits.

5.5 Numerical Results

We have investigated the tail asymptotics of cycle times and waiting times in Sections 5.2 and 5.3. In this section, we provide numerical experiments to study the convergence behavior of the transient cycle times and waiting times for closed tandem queues with communication blocking and manufacturing blocking rules.

We consider ten systems as given in Table 1. More specifically, Systems 1 through 5 have the same number of stations (i.e., $K = 5$) and the same system capacity (i.e., $M = 20$) but different number of customers in the system. On the other hand, Systems 6 through 10 have the same number of stations (i.e., $K = 10$) and the same system capacity (i.e., $M = 20$) but different number of customers in the system.

We assume that service times at all stations for all ten systems have Pareto distribution with parameter 1 (i.e., $\overline{B}_k(x) = x^{-1}$ for all $k \in \{1, \ldots, K\}$). Then, $c_k = 1$ for all $k \in \{1, \ldots, K\}$. The tail asymptotics of the transient cycle times and waiting times are independent of n as long as $n \geq N$ as discussed in Section 5.2.2. Thus, we consider the system time which is the cycle time corresponding to the last station and waiting times at the first station and the last station of the 20th customer in all ten systems. In particular, for each value of x, we first approximate the tail probabilities

48

Table 1: Description of ten systems that we consider

System	K	M_k for all $k \in \{1, \ldots, K\}$	M	N
System 1	5	4	20	2
System 2	5	4	20	5
System 3	5	4	20	10
System 4	5	4	20	15
System 5	5	4	20	19
System 6	10	2	20	2
System 7	10	2	20	5
System 8	10	2	20	10
System 9	10	2	20	15
System 10	10	2	20	19

of system times and waiting times using the tail asymptotics of system times and waiting times (as given in Propositions 5.2.4 and 5.2.5, respectively) and compare them with the tail probabilities obtained from simulation analysis. In our simulation study, for each value of x we run 41 batches of 10,000 replications and compute the average and 95% confidence interval of the corresponding tail probability. However, for purposes of clarity, we do not present the confidence intervals in Figures 11 to 40.

Figures 11 through 25 display the tail asymptotics and simulation results for the transient system times and waiting times for all ten systems with communication blocking rule. On the other hand, Figures 26 through 40 present numerical results for all ten systems with manufacturing blocking rule.

First of all, we focus on numerical results on systems with communication blocking. Figure 11 displays the tail probabilities of system times of the 20th customer in Systems 1, 2, 3, 4, and 5 which have the same number of stations and the same capacity but different number of customers in the system. This figure illustrates that the tail asymptotics could be used to approximate the tail probabilities of system times even when x is moderately large except System 5. In particular, when the total number of customers in system is small (i.e., System 1), the convergence of the
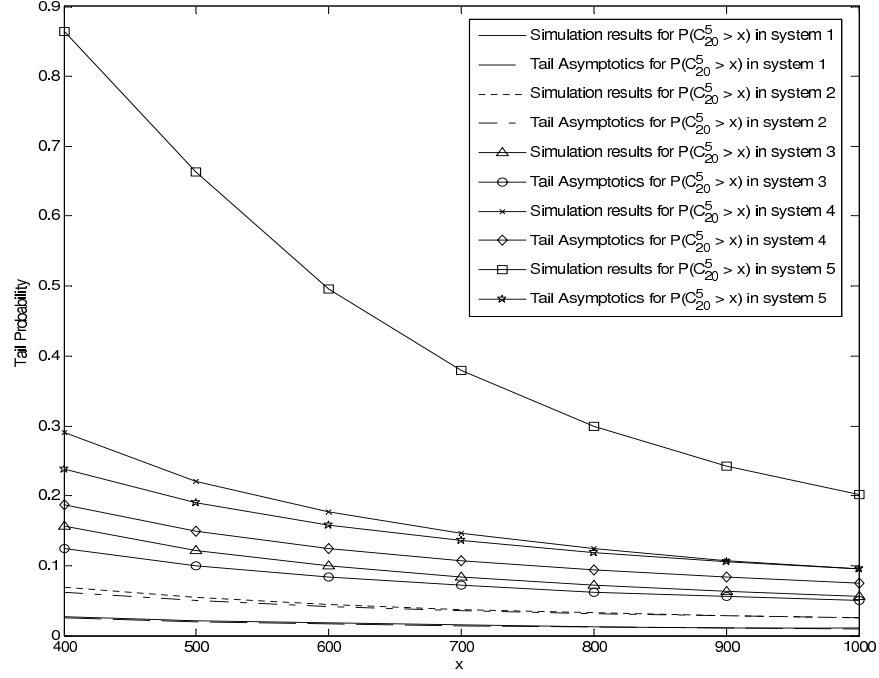
Figure 11: Simulation results and the tail asymptotics of system times for Systems 1, 2, 3, 4, and 5 with communication blocking

tail asymptotics to the actual tail probability is fast. Figures 12 and 13 present the tail probabilities of waiting times at stations 1 and 5 (last station) for Systems 1, 2, 3, 4, and 5. As these figures illustrate, the tail asymptotics could provide a good approximation for the tail probabilities of waiting times even when x is moderately large except System 5. Especially, the convergence of the tail asymptotics to the tail probabilities is fast for the waiting times when the number of customers in system is small. However, the convergence for System 4 is slightly better than System 3. Note that the tail asymptotics of waiting times at stations 1 and 5 for the systems with the same number of customers are same since the buffer size of all stations is the same. We have similar observations for Systems 6, 7, 8, 9, and 10 which have the same number of stations and capacity but different number of customers in the system. Figures 14 through 16 present the tail asymptotics for system times and waiting times (at first station and last station) for Systems 6, 7, 8, 9, and 10.
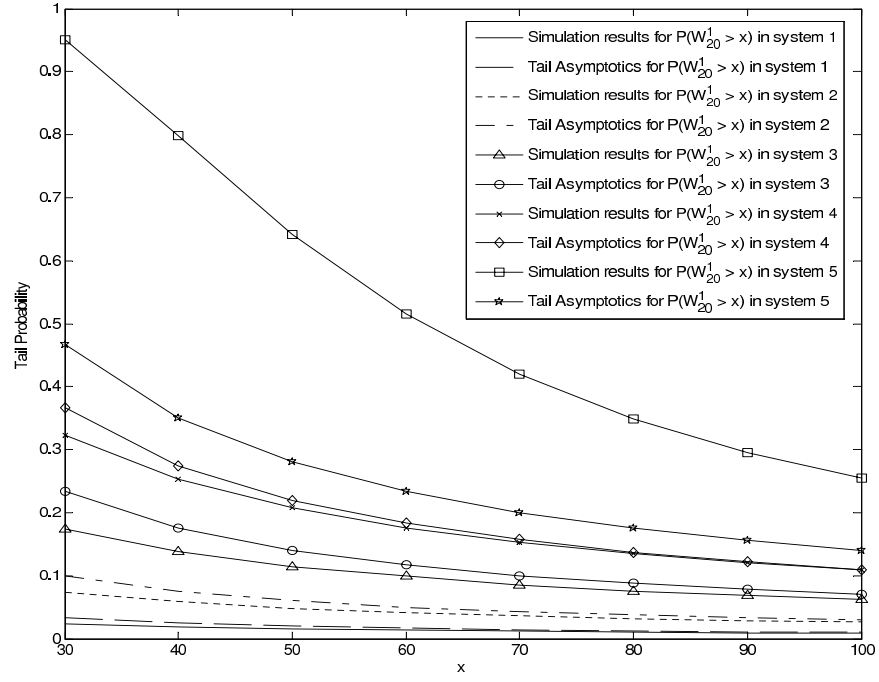
50

Figure 12: Simulation results and the tail asymptotics of waiting times at station 1 for Systems 1, 2, 3, 4, and 5 with communication blocking
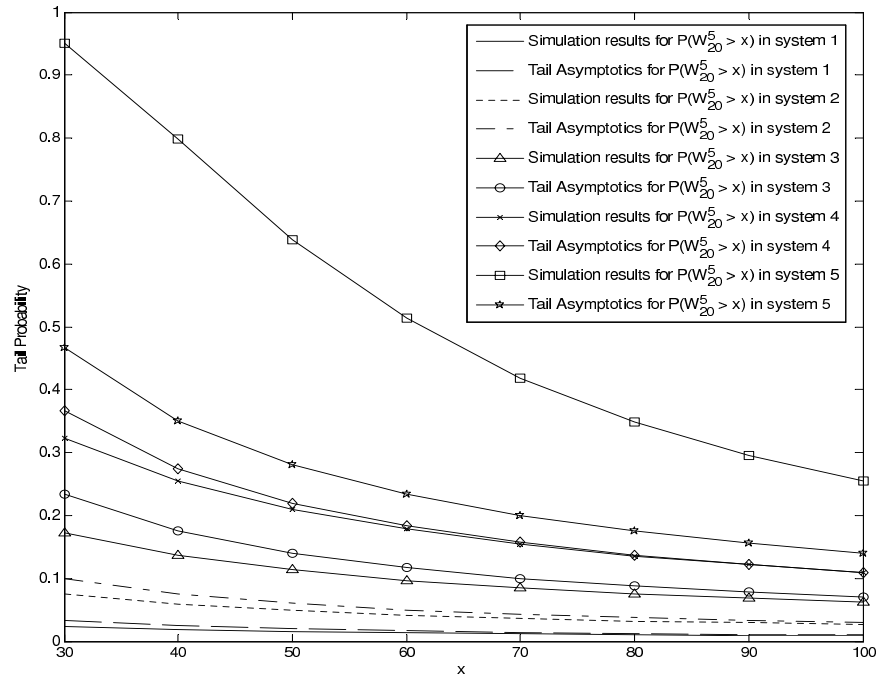

Figure 13: Simulation results and the tail asymptotics of waiting times at station 5 for Systems 1, 2, 3, 4, and 5 with communication blocking
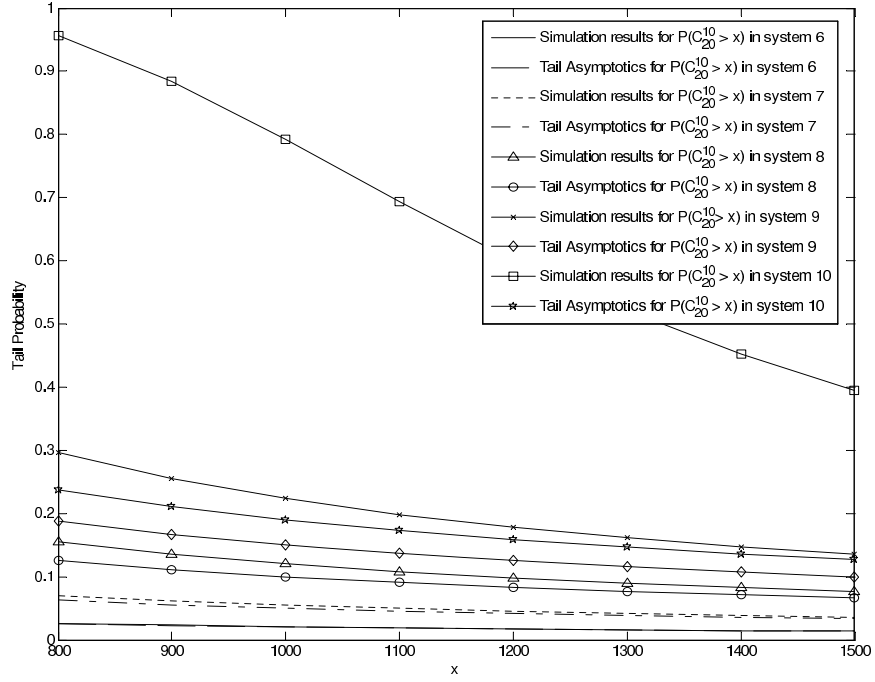
Figure 14: Simulation results and the tail asymptotics of system times for Systems 6, 7, 8, 9, and 10 with communication blocking
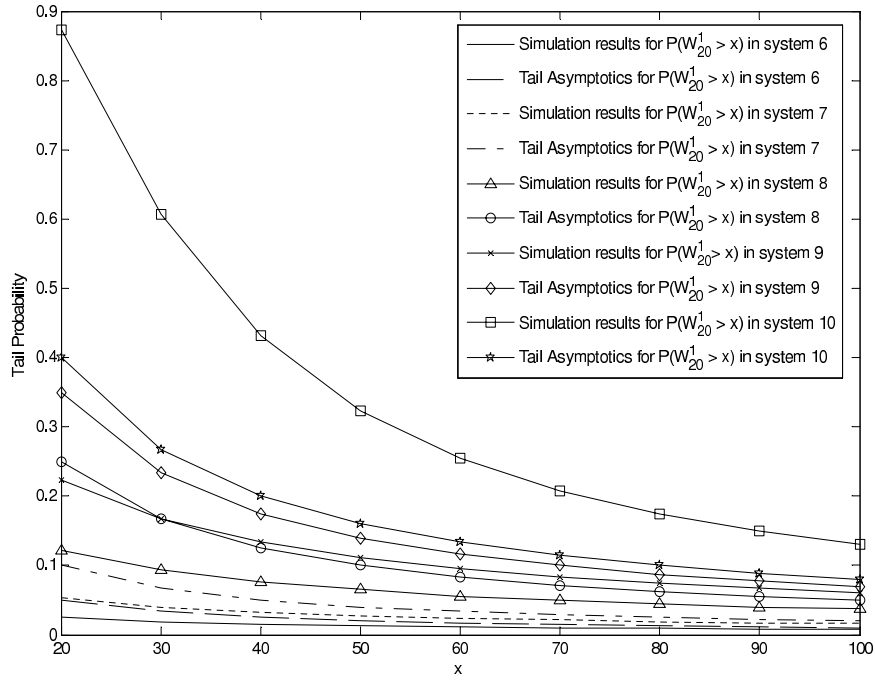

Figure 15: Simulation results and the tail asymptotics of waiting times at station 1 for Systems 6, 7, 8, 9, and 10 with communication blocking

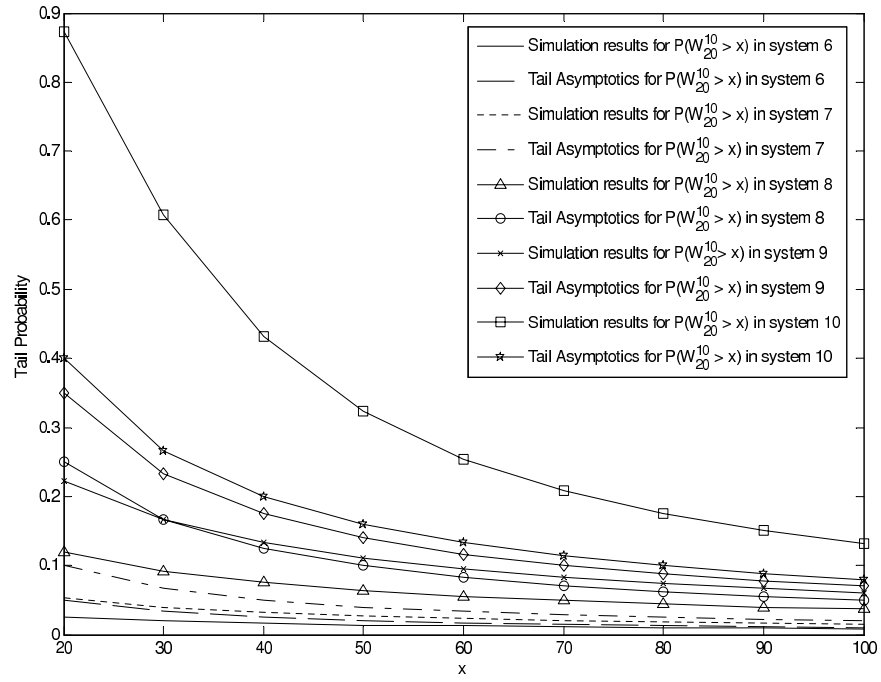**Figure 16:** Simulation results and the tail asymptotics of waiting times at station 10 for Systems 6, 7, 8, 9, and 10 with communication blocking
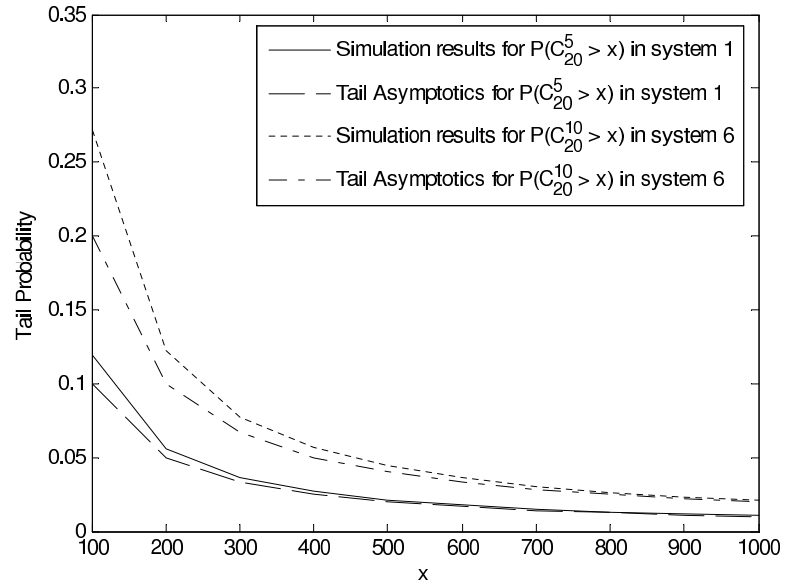

Figure 17: Simulation results and the tail asymptotics of system times for Systems 1 and 6 with communication blocking
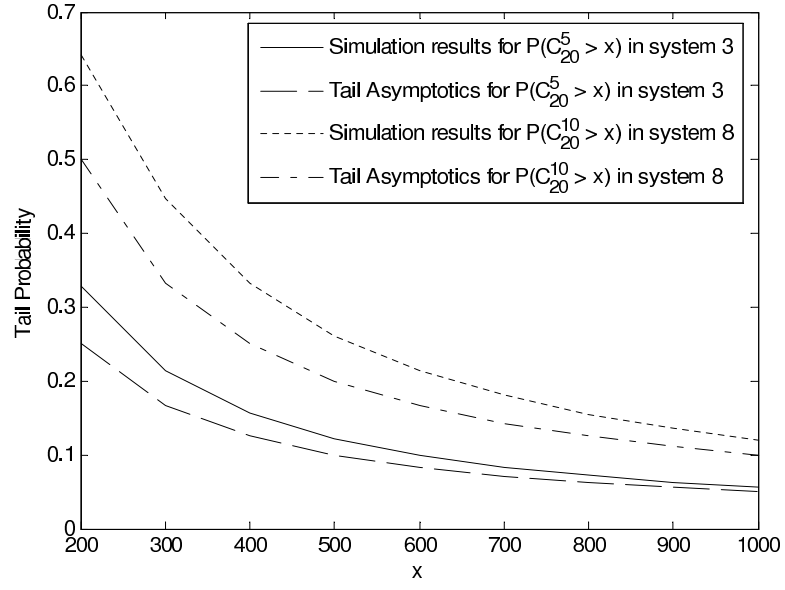
53

Figure 18: Simulation results and the tail asymptotics of system times for Systems 3 and 8 with communication blocking
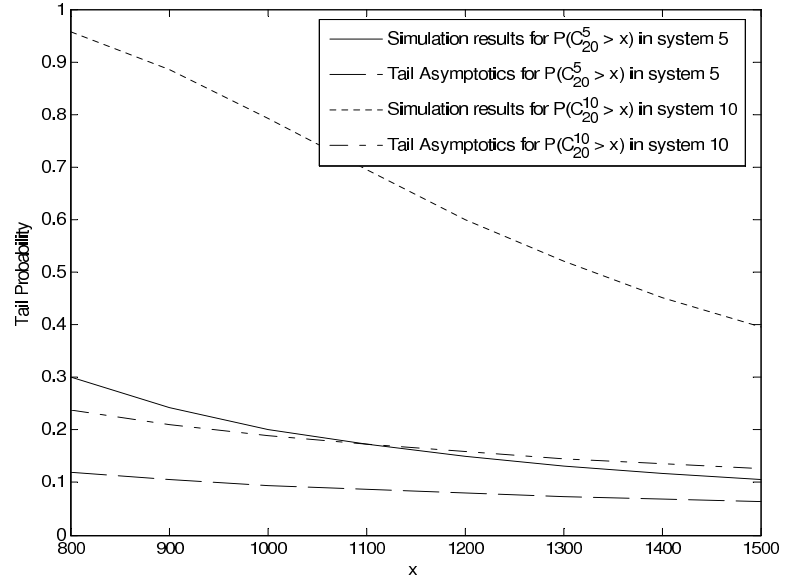

Figure 19: Simulation results and the tail asymptotics of system times for Systems 5 and 10 with communication blocking

54

Figure 20: Simulation results and the tail asymptotics of waiting times of the first station for Systems 1 and 6 with communication blocking

Figure 21: Simulation results and the tail asymptotics of waiting times of the first station for Systems 3 and 8 with communication blocking

Figure 22: Simulation results and the tail asymptotics of waiting times of the first station for Systems 5 and 10 with communication blocking

Figure 23: Simulation results and the tail asymptotics of waiting times of the last station for Systems 1 and 6 with communication blocking
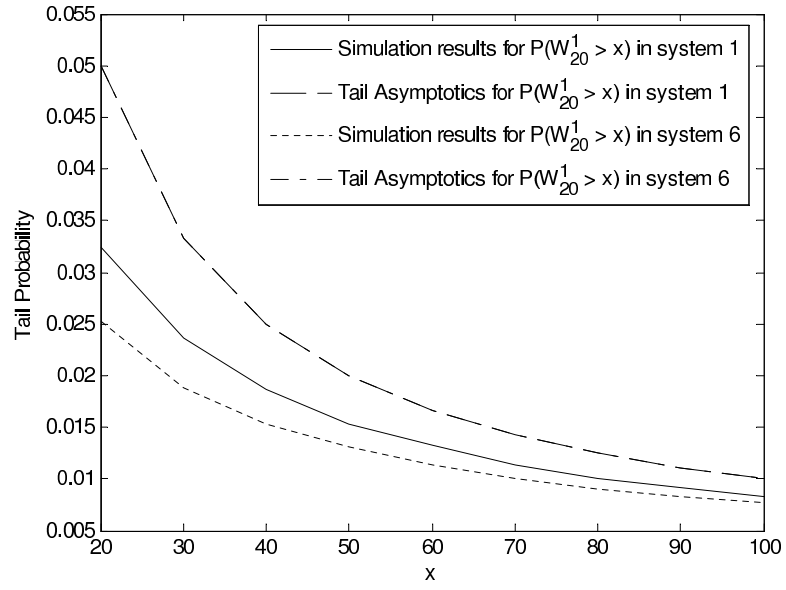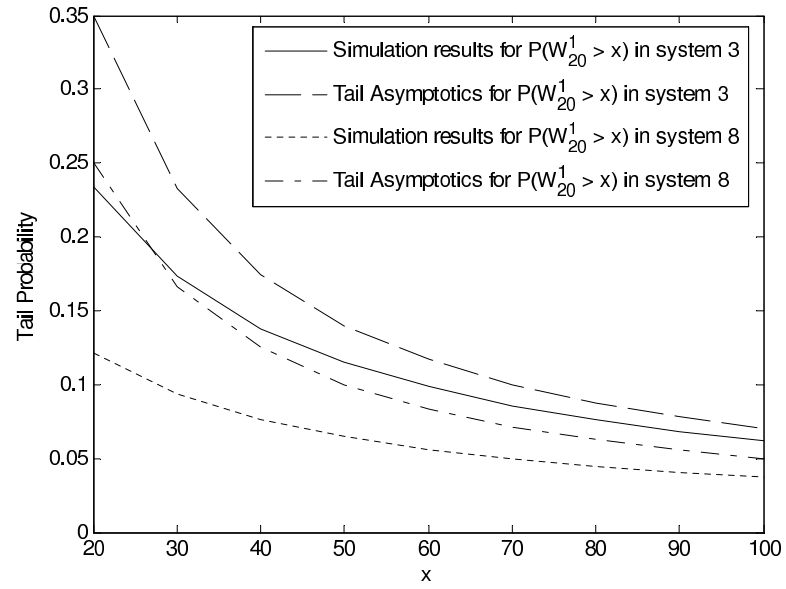
56

Figure 24: Simulation results and the tail asymptotics of waiting times of the last station for Systems 3 and 8 with communication blocking
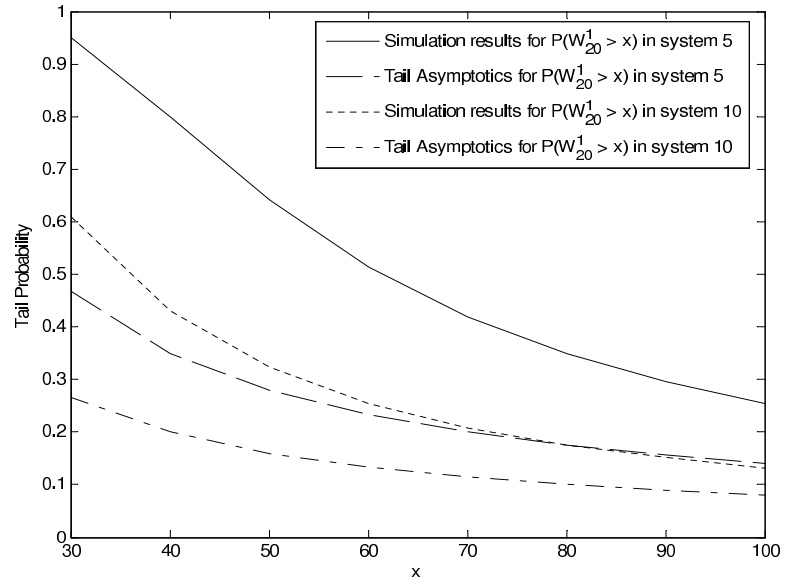

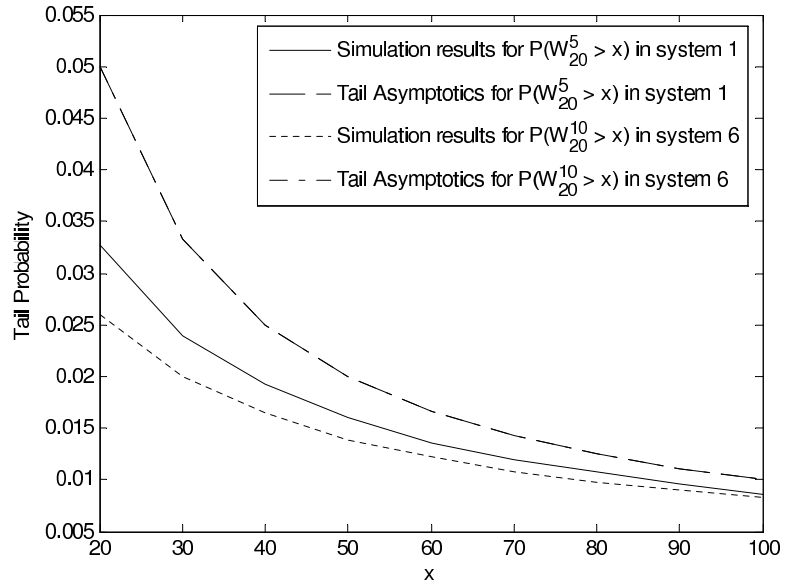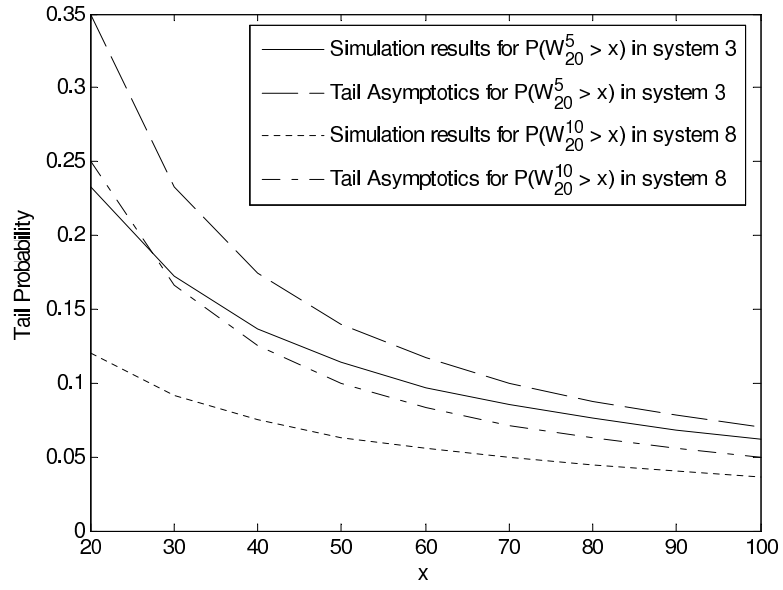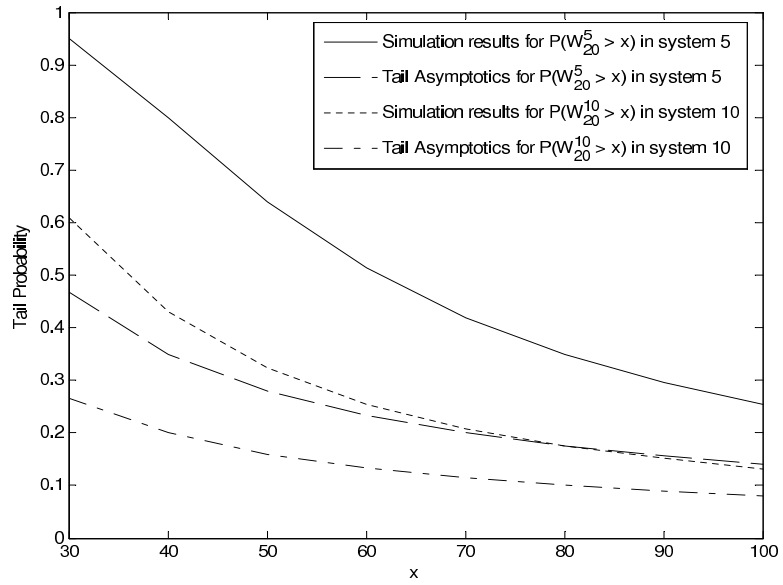**Figure 25:** Simulation results and the tail asymptotics of waiting times of the last station for Systems 5 and 10 with communication blocking

Figures 17 through 25 illustrate the convergence behavior of the tail asymptotics of the system times and waiting times in two systems which have the same system capacity and the same number of customers in the system but different number of stations. Note that the convergence of the tail asymptotics is fast for systems with the small number of stations. However, from figures 22 and 25 (i.e., when $N = 19$), the tail asymptotics of waiting times for System 5 (i.e., $K = 5$) converge slower than System 10 (i.e., $K = 10$) since waiting times depend on the buffer size as well. Note that in Figure 20 and 23, the tail asymptotics of the waiting times at first and last stations for Systems 1 and 6 are the same because both systems have the same number of customers and no blocking. As we have observed above, in all these cases the tail asymptotics could provide a good approximation for the tail probability of the system times and waiting times as x increases from medium to large values except Systems 5 and 10.

Figures 26 through 40 present the tail asymptotics and simulation results for the transient system times and waiting times for all ten systems with manufacturing blocking rule. These figures show that the numerical results for communication blocking and manufacturing blocking for Systems 1, 2, 3, 6, 7, and 8 are very similar. However, the convergence for Systems 4, 5, 9, and 10 (i.e., $N = 15$ and $N = 19$) with manufacturing blocking is faster than communication blocking.

Figure 26: Simulation results and the tail asymptotics of system times for Systems 1, 2, 3, 4, and 5 with manufacturing blocking

Figure 27: Simulation results and the tail asymptotics of waiting times at station 1 for Systems 1, 2, 3, 4, and 5 with manufacturing blocking

Figure 28: Simulation results and the tail asymptotics of waiting times at station 5 for Systems 1, 2, 3, 4, and 5 with manufacturing blocking
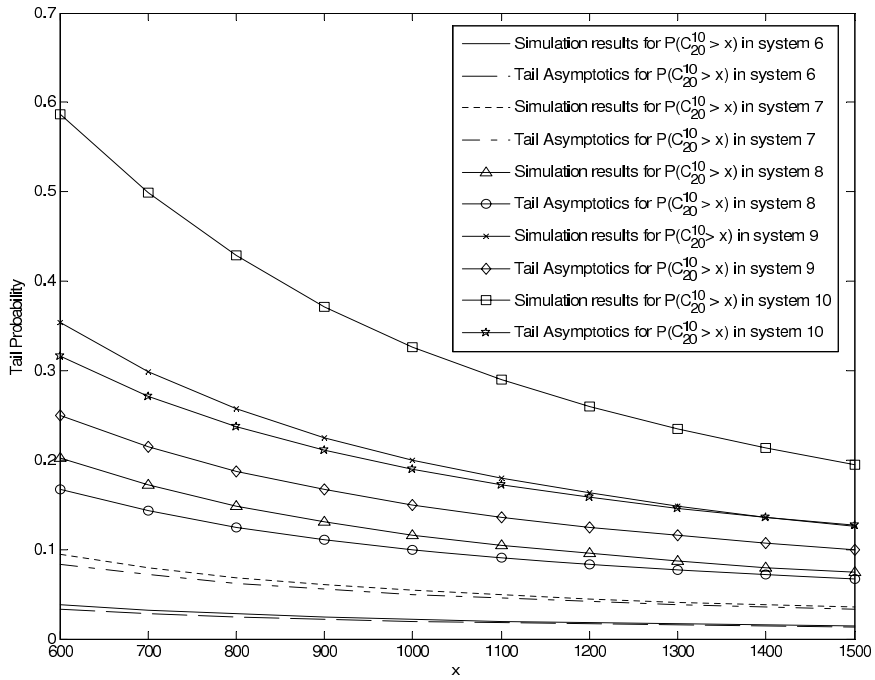
Figure 29: Simulation results and the tail asymptotics of system times for Systems 6, 7, 8, 9, and 10 with manufacturing blocking
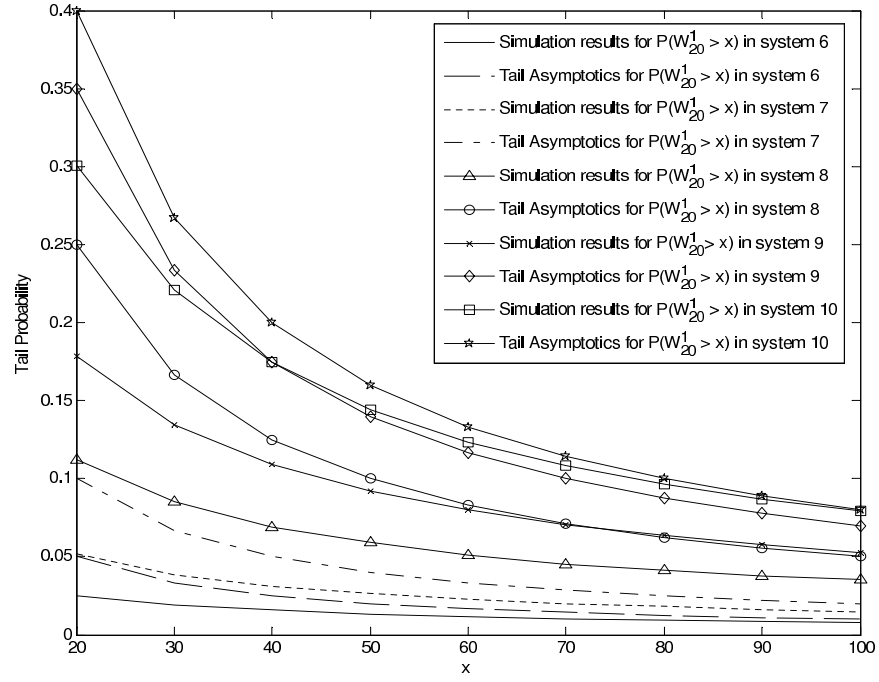
Figure 30: Simulation results and the tail asymptotics of waiting times at station 1 for Systems 6, 7, 8, 9, and 10 with manufacturing blocking
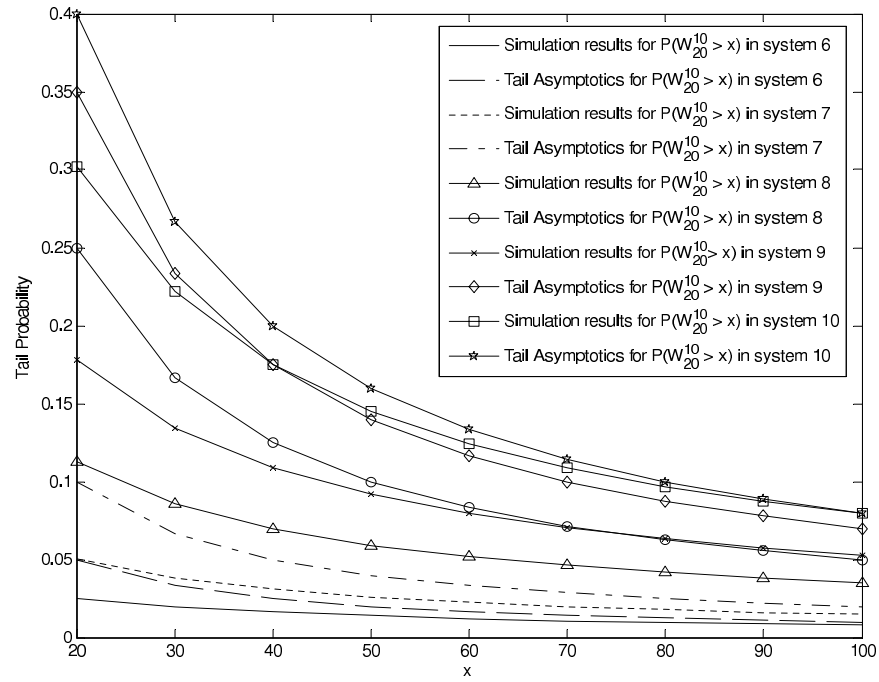
Figure 31: Simulation results and the tail asymptotics of waiting times at station 10 for Systems 6, 7, 8, 9, and 10 with manufacturing blocking
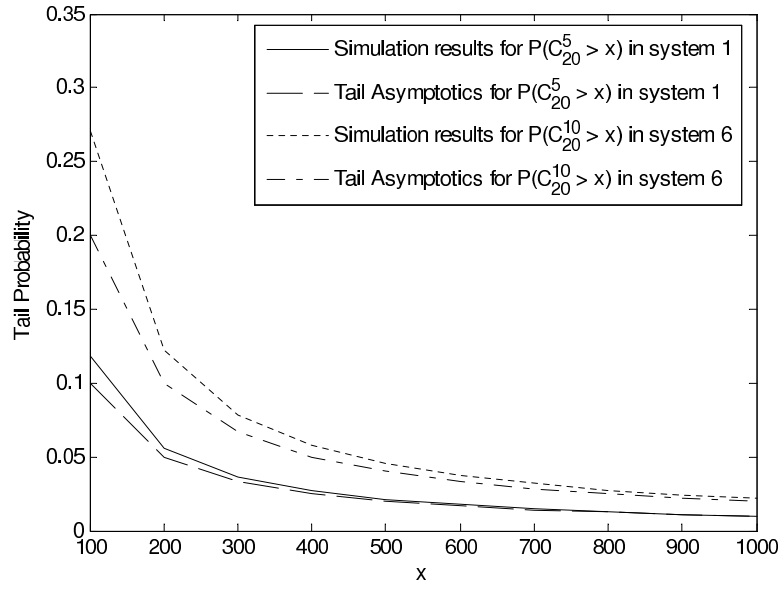
Figure 32: Simulation results and the tail asymptotics of system times for Systems 1 and 6 with manufacturing blocking
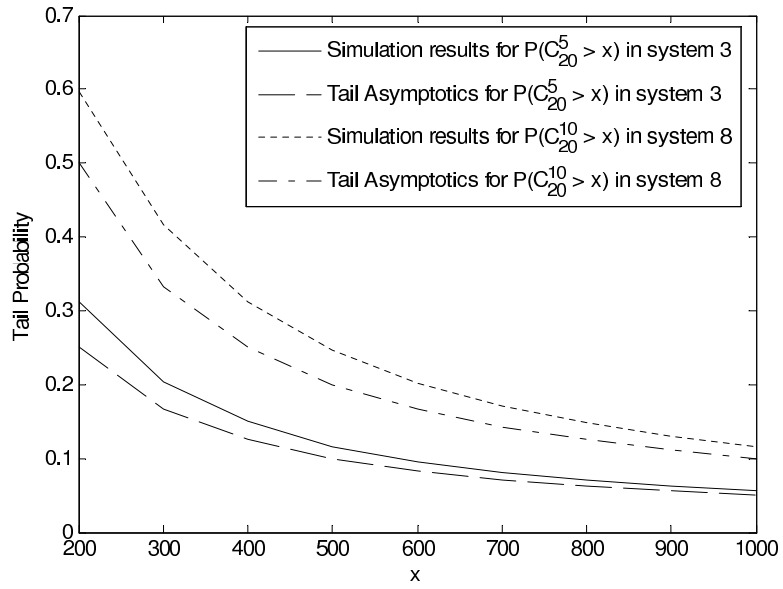
Figure 33: Simulation results and the tail asymptotics of system times for Systems 3 and 8 with manufacturing blocking
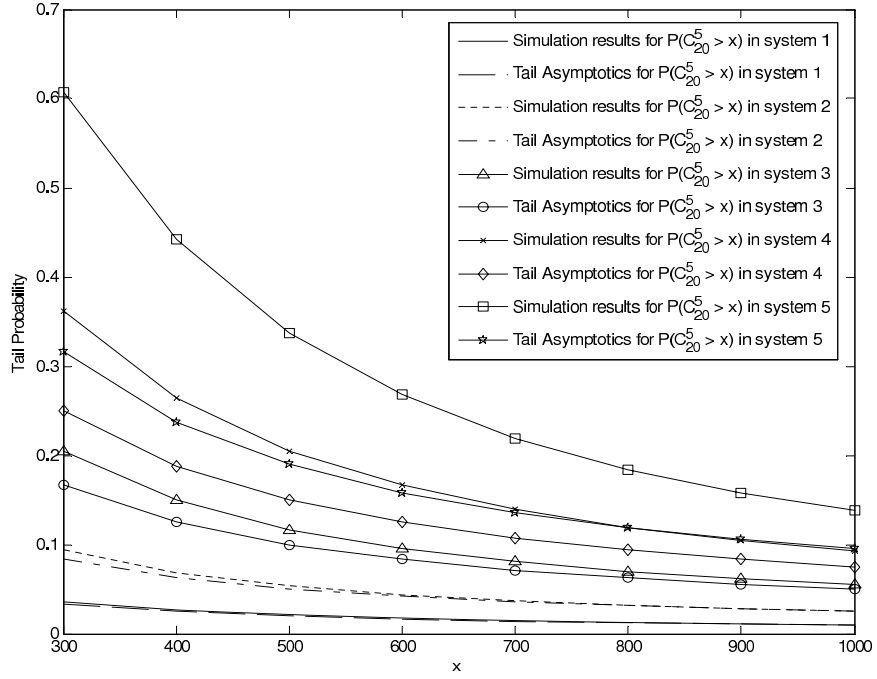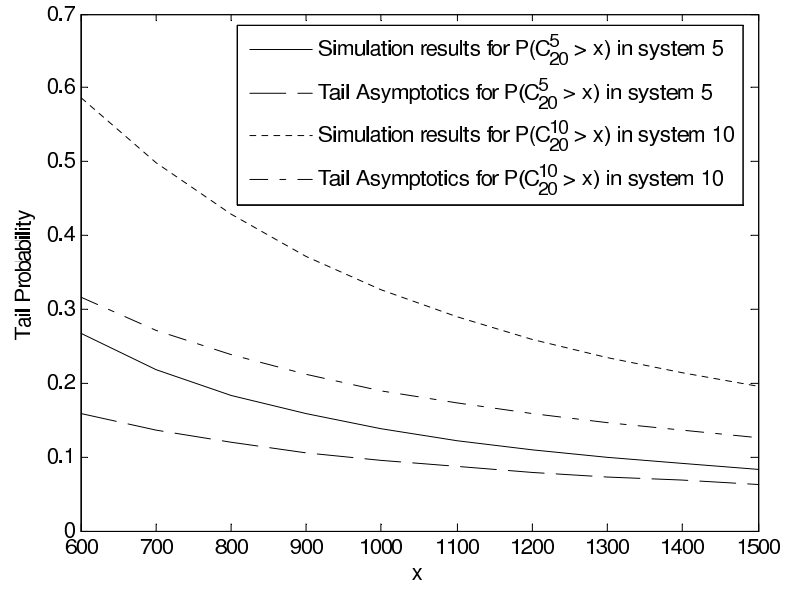
Figure 34: Simulation results and the tail asymptotics of system times for Systems 5 and 10 with manufacturing blocking
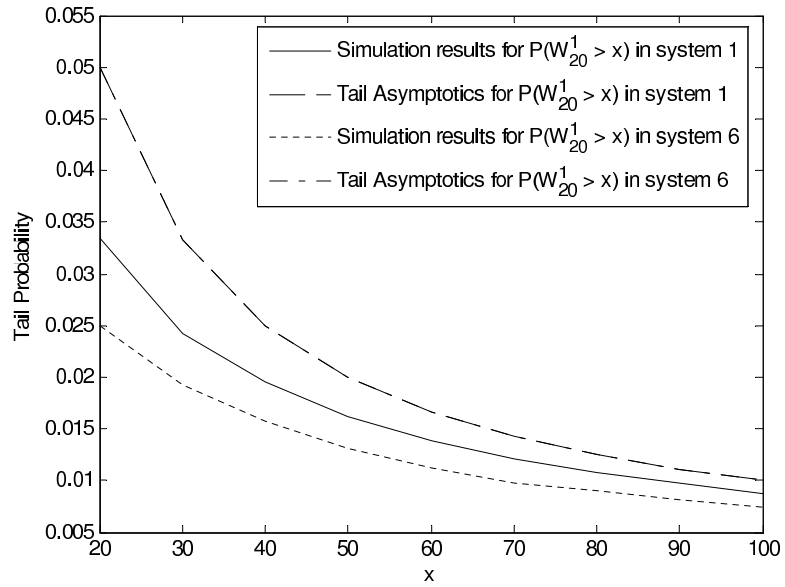
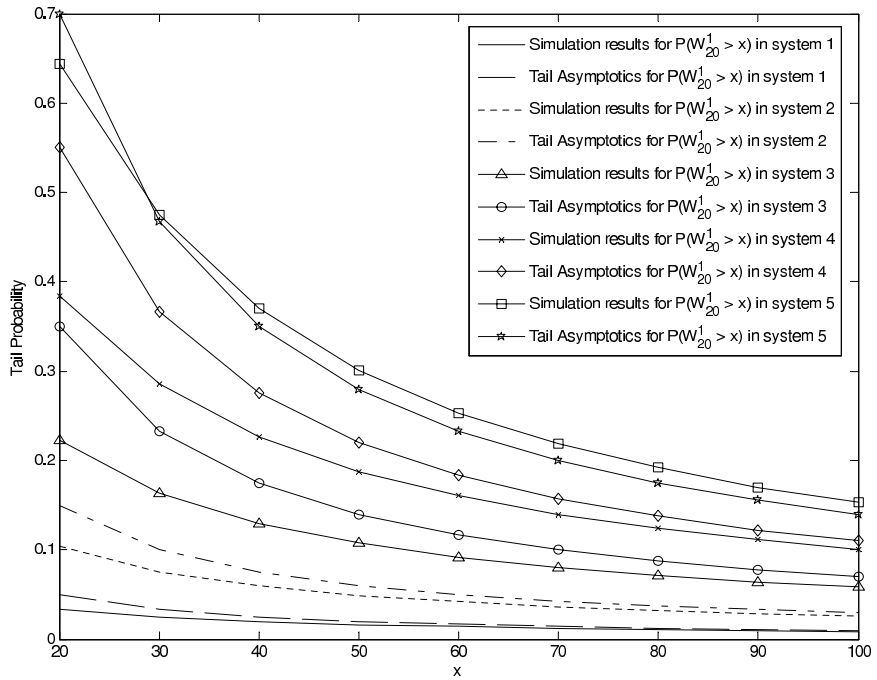**Figure 35:** Simulation results and the tail asymptotics of waiting times of the first station for Systems 1 and 6 with manufacturing blocking

63
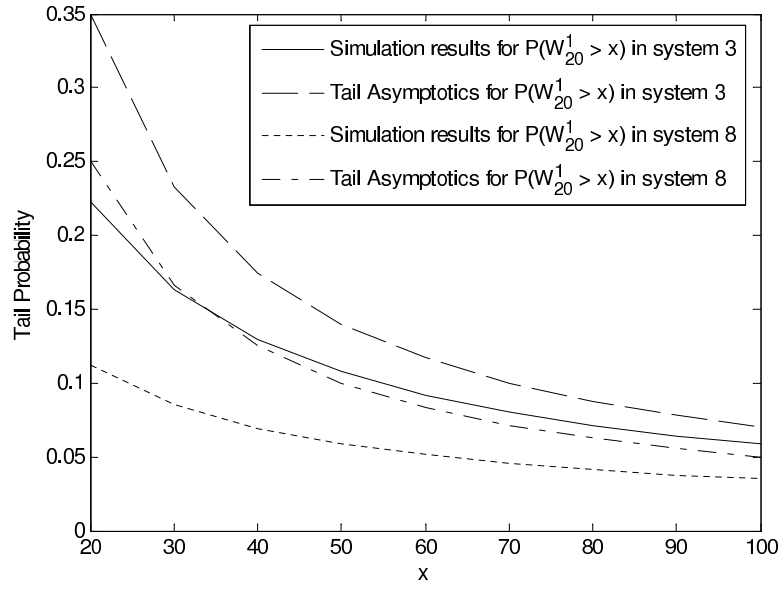
Figure 36: Simulation results and the tail asymptotics of waiting times of the first station for Systems 3 and 8 with manufacturing blocking
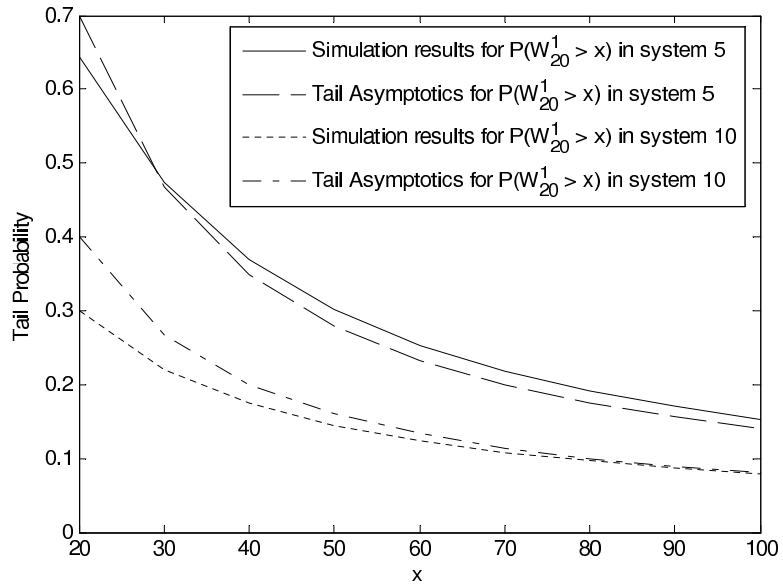
Figure 37: Simulation results and the tail asymptotics of waiting times of the first station for Systems 5 and 10 with manufacturing blocking
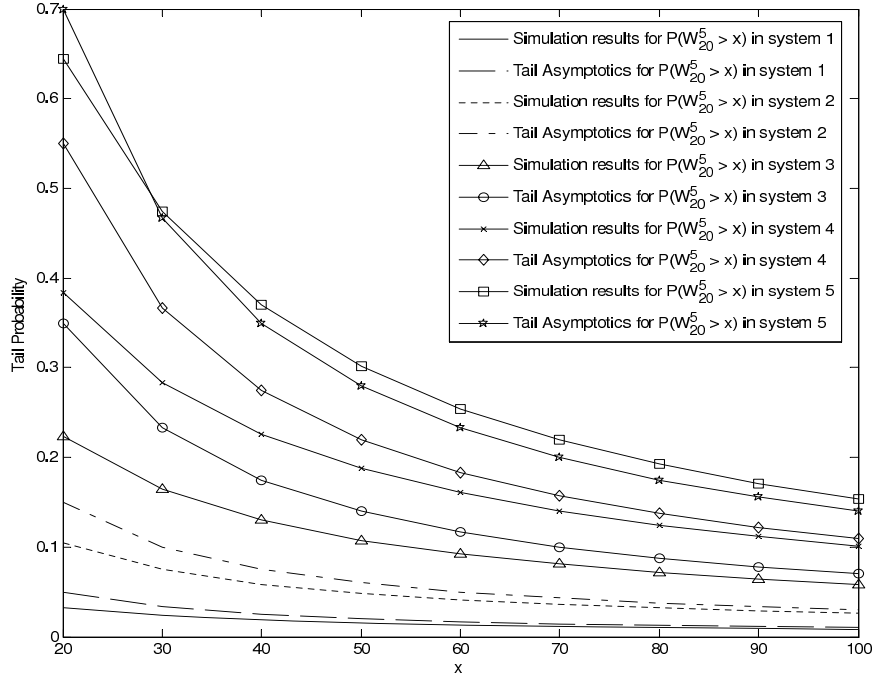
64
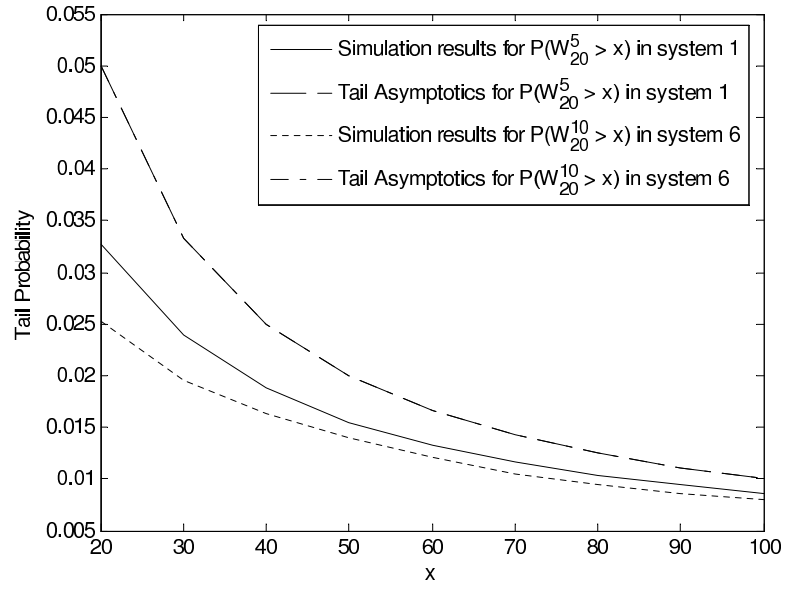
Figure 38: Simulation results and the tail asymptotics of waiting times of the last station for Systems 1 and 6 with manufacturing blocking
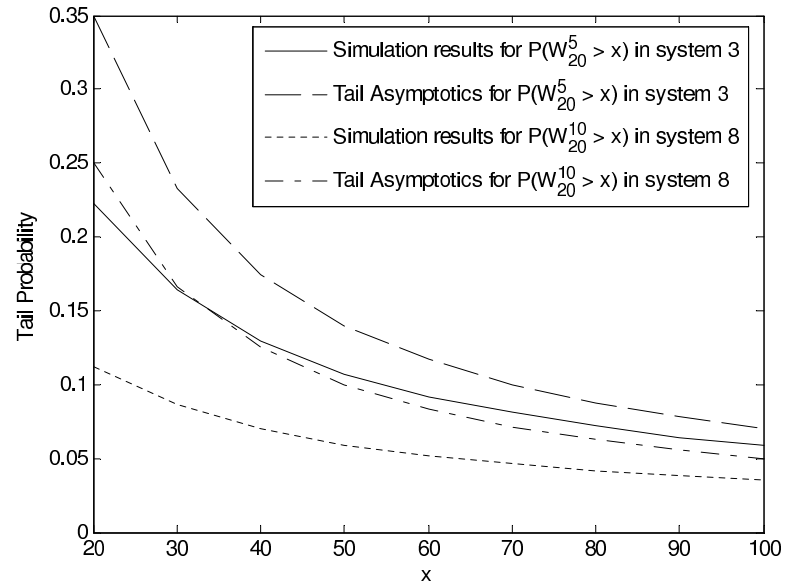
Figure 39: Simulation results and the tail asymptotics of waiting times of the last station for Systems 3 and 8 with manufacturing blocking

65
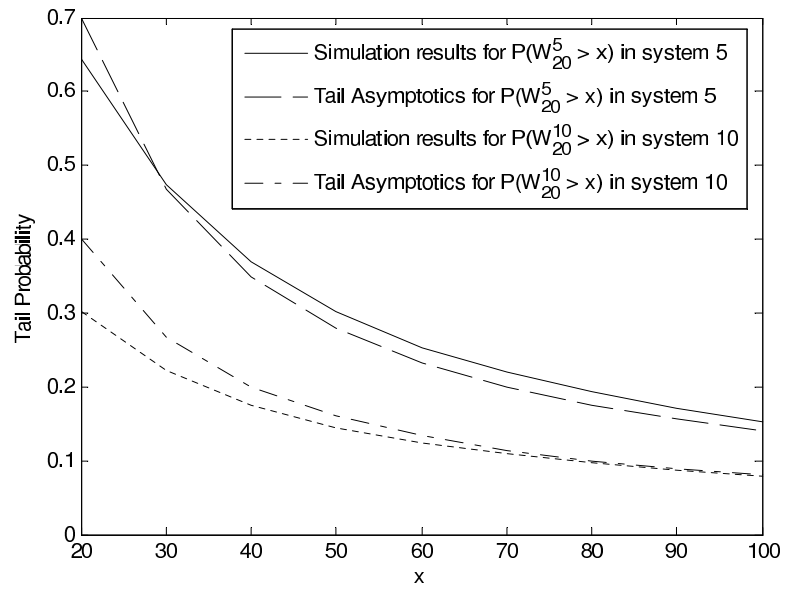
Figure 40: Simulation results and the tail asymptotics of waiting times of the last station for Systems 5 and 10 with manufacturing blocking

CHAPTER VI

TANDEM QUEUES WITH SUBEXPONENTIAL SERVICE TIMES AND FINITE BUFFERS

6.1 Introduction

We consider open tandem queues with subexponential service time distributions and blocking which is caused by finite buffer capacities between stations. More specifically, we focus on a K-station ($K \geq 2$) tandem network of single-server stations with an infinite number of customers in front of the first station and infinite room for finished customers after last station but finite buffers between stations k and $k+1$ for $k = 1, \ldots, K-1$ as shown in Figure 41. We analyze this system under communication blocking and manufacturing blocking schemes. Our objective is to derive the tail asymptotics of transient and stationary response times and waiting times in these networks.

As defined in Chapter 5, let M_{k+1} be the size of the buffer between station k and $k + 1$ for $k = 1, \ldots, K - 1$ including the customer being served at station $k + 1$. As mentioned above we assume that $0 < M_{k+1} < \infty$ for $k = 1, \ldots, K - 1$ and for notational convenience we set $M_1 = 1$ to denote the buffer capacity of the first station. As in Chapter 5, let N_k be the initial number of customers at station $k \in \{1, \ldots, K\}$ (including those waiting in the buffer). Clearly, $0 \leq N_k \leq M_k$. Since there are infinite number of customers in front of station 1 and $M_1 = 1$, without loss of generality we set $N_1 = 0$. We assume that all stations are idle at time 0 and if there is a customer at a station, the service on that customer has not started before time 0. Finally, as in Chapter 5, $H_k = M_k - N_k$ denotes the number of empty spaces in the buffer of station $k \in \{1, \ldots, K\}$ at time 0. Clearly, $0 \leq H_k \leq M_k$ for all $k \in \{2, \ldots, K\}$ and $H_1 = 1$.

67

The service discipline at all stations is assumed to be First Come First Served (FCFS). Service times at station $k \in \{1, \ldots, K\}$ are independent and identically distributed random variables $\{B_n^k\}$ with distribution function $B_k(\cdot)$. The sequence of service times at each station is independent of the service times at the other stations. Furthermore, we assume that there exists a subexponential distribution $F(\cdot)$ ($F \in \mathcal{S}$) and there exist constants $c_k \in [0, \infty)$ with $\sum_{k=1}^{K} c_k > 0$ such that for all $k \in \{1, \ldots, K\}$

$$\lim_{x \to \infty} \frac{\overline{B}_k(x)}{\overline{F}(x)} = c_k. \tag{24}$$
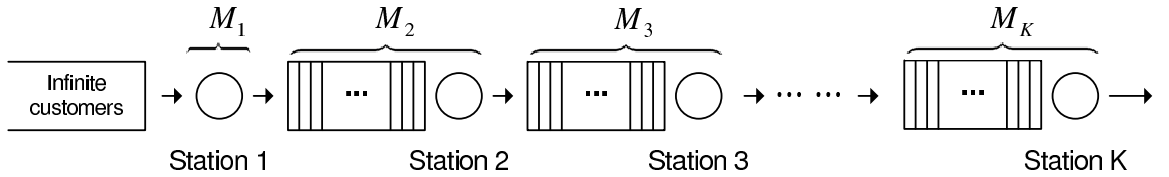
Figure 41: A finite buffer tandem queueing network with K stations.

The remainder of this chapter is organized as follows. Tandem queues with communication blocking are discussed in Section 6.2. More specifically, Section 6.2.1 gives some preliminary results and Section 6.2.2 provides the asymptotics of the nth response time and n^{th} waiting time at station k for all $k \in \{1, \ldots, K\}$. Similarly, in Section 6.3, we focus on the tail asymptotics of transient response times and waiting times in tandem queues with manufacturing blocking. Section 6.4 concentrates on stationary response times and waiting times. Numerical experiments investigating the convergence of these tail probabilities to their asymptotic counter parts are provided in Section 6.5.

6.2 Communication Blocking

In this section, we consider the tandem network of Section 6.1 when it is operating under communication blocking and derive the tail asymptotics for transient response times and waiting times.

6.2.1 Preliminaries

We first derive a recursive expression for the departure times. For notational convenience, as in Chapter 4 and 5, define

$$\mathcal{N}_{k,u} = \sum_{j=k}^{u} N_j,$$

with the convention that summation over an empty set is zero. However, note that $\mathcal{N}_{k,u}$ is different from one used in Chapter 4 and 5. Unlike Chapter 4 and 5, in this chapter, $\mathcal{N}_{k,u}$ is the total number of initial customers from station k to station u in the direction of customer flow for $u \geq k$ and $\mathcal{N}_{k,k} = N_k$. Similarly, as in Chapter 5, define

$$\mathcal{H}_{k,u} = \sum_{j=k}^{u} H_j,$$

with the convention that summation over an empty set is zero. Hence, $\mathcal{H}_{k,u}$ is the total number of initial empty spaces in the buffers of the stations from station k to station u in the direction of customer flow for $u \geq k$ and $\mathcal{H}_{k,k} = H_k$. We again use X_n^k to denote the departure time of the nth customer from station $k \in \{1, \ldots, K\}$. Then, we have the following expression.

Proposition 6.2.1 *For all $k \in \{1, \ldots, K\}$ and $n \geq 1$,*

$$X_n^k = \max\{X_{n-1}^k + B_n^k,\ X_{n-N_k}^{k-1} + B_n^k,\ X_{n-H_{k+1}}^{k+1} + B_n^k\} \tag{25}$$

with the convention that $X_n^0 = 0$, $X_n^{K+1} = 0$ for all n and $X_n^k = 0$ for all $n \leq 0$ and all $k \in \{1, \ldots, K\}$.

Proof follows immediately from the observation that the nth customer starts the server at station k at time $\max\{X_{n-1}^k,\ X_{n-N_k}^{k-1},\ X_{n-H_{k+1}}^{k+1}\}$. $\qquad\square$

In the proof of our main results, we will make use of the following upper and lower bounds on X_n^k for all $k \in \{1, \ldots, K\}$.

Proposition 6.2.2 *For all $k \in \{1, \ldots, K\}$ and $n \geq 1$,*

$$X_n^k \leq \sum_{u=1}^{k} \sum_{r=1}^{n-\mathcal{N}_{u+1,k}} B_r^u + \sum_{u=k+1}^{K} \sum_{r=1}^{n-\mathcal{H}_{k+1,u}} B_r^u$$

with the convention that the summation over an empty set is zero.

Proof follows from summing up all the service times that appear in (25). □

Proposition 6.2.3 *For all $k \in \{1, \ldots, K\}$ and $n \geq 1$,*

$$X_n^k \geq \max\{ \max_{u=1,\ldots,k} \sum_{r=1}^{n-\mathcal{N}_{u+1,k}} B_r^u, \max_{u=k+1,\ldots,K} \sum_{r=1}^{n-\mathcal{H}_{k+1,u}} B_r^u \}$$

with the convention that the summation over an empty set is equal to zero and the maximization over an empty set is equal to $-\infty$.

Proof We obtain lower bounds on all three terms that appear in the maximization operation of (25). It follows from equation (25) that $X_n^u \geq X_{n-1}^u + B_n^u$ for all $u \in \{1, ..., K\}$ and for all $n \geq 1$. Then by recursive substitution, we have

$$X_{n-1}^k + B_n^k \;\geq\; \sum_{r=1}^{n} B_r^k. \tag{26}$$

From (25), we have $X_n^u \geq \max\{X_{n-1}^u + B_n^u, \; X_{n-N_u}^{u-1} + B_n^u\} \geq \max\{X_{n-1}^u + B_n^u, \; X_{n-N_u}^{u-1}\}$ for all $u \in \{1, ..., K\}$ and $n \geq 1$. Using these inequalities recursively, we obtain

$$X_{n-N_k}^{k-1} + B_n^k \;\geq\; \max\{X_{n-N_k-1}^{k-1} + B_{n-N_k}^{k-1}, \; X_{n-\mathcal{N}_{k-1,k}}^{k-2}\}$$

$$\geq\; \max\{X_{n-\mathcal{N}_{k,k}-1}^{k-1} + B_{n-\mathcal{N}_{k,k}}^{k-1}, \; X_{n-\mathcal{N}_{k-1,k}-1}^{k-2} + B_{n-\mathcal{N}_{k-1,k}}^{k-2}, \; X_{n-\mathcal{N}_{k-2,k}}^{k-3}\}$$

$$\vdots$$

$$\geq\; \max_{u=1,\ldots,k-1}\{X_{n-\mathcal{N}_{u+1,k}-1}^{u} + B_{n-\mathcal{N}_{u+1,k}}^{u}\}$$

$$\geq\; \max_{u=1,\ldots,k-1}\{ \sum_{r=1}^{n-\mathcal{N}_{u+1,k}} B_r^u \} \tag{27}$$

70

where the last inequality follows from (26). Similarly, from (25), $X_n^u \geq \max\{X_{n-1}^u + B_n^u, X_{n-H_{u+1}}^{u+1} + B_n^u\} \geq \max\{X_{n-1}^u + B_n^u, X_{n-H_{u+1}}^{u+1}\}$. Then we have

$$X_{n-H_{k+1}}^{k+1} + B_n^k \geq \max\{X_{n-H_{k+1}-1}^{k+1} + B_{n-H_{k+1}}^{k+1}, X_{n-\mathcal{H}_{k+1,k+2}}^{k+2}\}$$

$$\geq \max\{X_{n-\mathcal{H}_{k+1,k+1}-1}^{k+1} + B_{n-\mathcal{H}_{k+1,k+1}}^{k+1}, X_{n-\mathcal{H}_{k+1,k+2}-1}^{k+2} + B_{n-\mathcal{H}_{k+1,k+2}}^{k+2}, X_{n-\mathcal{H}_{k+1,k+3}}^{k+3}\}$$

$$\vdots$$

$$\geq \max_{u=k+1,\ldots,K}\{X_{n-\mathcal{H}_{k+1,u}-1}^u + B_{n-\mathcal{H}_{k+1,u}}^u\}$$

$$\geq \max_{u=k+1,\ldots,K}\{\sum_{r=1}^{n-\mathcal{H}_{k+1,u}} B_r^u\}. \tag{28}$$

Putting (26), (27), and (28) together, for all $k \in \{1,\ldots,K\}$, we have

$$X_n^k \geq \max\{\sum_{r=1}^{n} B_r^k, \max_{u=1,\ldots,k-1}\{\sum_{r=1}^{n-\mathcal{N}_{u+1,k}} B_r^u\}, \max_{u=k+1,\ldots,K}\{\sum_{r=1}^{n-\mathcal{H}_{k+1,u}} B_r^u\}\}$$

$$= \max\{\max_{u=1,\ldots,k}\sum_{r=1}^{n-\mathcal{N}_{u+1,k}} B_r^u, \max_{u=k+1,\ldots,K}\sum_{r=1}^{n-\mathcal{H}_{k+1,u}} B_r^u\}.$$

$\square$

6.2.2 Response Times and Waiting Times

Let R_n^k denote the response time of the nth customer at station k which is the time from his acceptance to station 1 to his departure from station k. Thus, the response time of the nth customer at station $k \in \{1,\ldots,K\}$ is computed as

$$R_n^k = X_n^k - X_{n-\mathcal{N}_{2,k}-1}^1. \tag{29}$$

Note that $X_{n-\mathcal{N}_{2,k}-1}^1$ is the time that the $(n-\mathcal{N}_{2,k})^{th}$ customer joins the server at station 1. For notational convenience, define

$$\mathcal{M}_{k,u} = \sum_{j=k}^{u} M_j,$$

with the convention that summation over an empty set is zero. Note that $\mathcal{M}_{k,u}$ is the total buffer capacity from station k to station u for $u \geq k$ and $\mathcal{M}_{k,k} = M_k$. As

71

in Chapter 5, let M denote the total capacity of the system. Thus, $M = \mathcal{M}_{1,K}$. The next proposition provides the tail asymptotics of R_n^k for all $n \geq \mathcal{M}_{1,k} + \mathcal{H}_{k+1,K}$.

Proposition 6.2.4 *For all $k \in \{1, \ldots, K\}$ and all $n \geq \mathcal{M}_{1,k} + \mathcal{H}_{k+1,K}$,*

$$\lim_{x \to \infty} \frac{\mathbb{P}(R_n^k > x)}{\overline{F}(x)} = \sum_{j=1}^{k} \mathcal{M}_{1,j}\, c_j + \mathcal{M}_{1,k} \sum_{j=k+1}^{K} c_j$$

where the convergence is uniform in n.

Proof Note that R_n^k will attain its largest value if all the stations (except station K) are blocked at the time that $(n - \mathcal{N}_{2,k})^{th}$ service starts at station 1. That is the $(n - \mathcal{N}_{2,k} - \mathcal{H}_{2,K})^{th}$ customer is still at station K when the $(n - \mathcal{N}_{2,k})^{th}$ service starts at station 1. Since $\mathcal{N}_{2,k} + \mathcal{H}_{2,j} = \mathcal{M}_{2,k} + \mathcal{H}_{k+1,j}$ for all $j \in \{k+1, \ldots, K\}$, we immediately obtain the following upper bound on R_n^k.

$$R_n^k \leq \sum_{j=1}^{k} \sum_{r=n-\mathcal{N}_{j+1,k}-\mathcal{M}_{2,j}}^{n-\mathcal{N}_{j+1,k}} B_r^j + \sum_{j=k+1}^{K} \sum_{r=n-\mathcal{M}_{2,k}-\mathcal{H}_{k+1,j}}^{n-\mathcal{H}_{k+1,j}} B_r^j.$$

Hence, from Corollary 3.0.1, for all $n \geq \mathcal{M}_{1,k} + \mathcal{H}_{k+1,K}$,

$$
\begin{aligned}
&\limsup_{x \to \infty} \frac{\mathbb{P}(R_n^k > x)}{\overline{F}(x)} \\
\leq\ &\limsup_{x \to \infty} \frac{\mathbb{P}(\sum_{j=1}^{k} \sum_{r=n-\mathcal{N}_{j+1,k}-\mathcal{M}_{2,j}}^{n-\mathcal{N}_{j+1,k}} B_r^j + \sum_{j=k+1}^{K} \sum_{r=n-\mathcal{M}_{2,k}-\mathcal{H}_{k+1,j}}^{n-\mathcal{H}_{k+1,j}} B_r^j > x)}{\overline{F}(x)} \\
=\ &\limsup_{x \to \infty} \frac{\mathbb{P}(\sum_{j=1}^{k} \sum_{r=1}^{\mathcal{M}_{2,j}+1} B_r^j + \sum_{j=k+1}^{K} \sum_{r=1}^{\mathcal{M}_{2,k}+1} B_r^j > x)}{\overline{F}(x)} \\
=\ &\sum_{j=1}^{k} \mathcal{M}_{1,j}\, c_j + \mathcal{M}_{1,k} \sum_{j=k+1}^{K} c_j.
\end{aligned}
\tag{30}
$$

We now provide a lower bound on R_n^k. From Propositions 6.2.2 and 6.2.3, we have

$$X_n^k - X_{n-\mathcal{N}_{2,k}-1}^1$$

$$\geq\ \max\left\{ \max_{j=1,\ldots,k} \sum_{r=1}^{n-\mathcal{N}_{j+1,k}} B_r^j,\ \max_{j=k+1,\ldots,K} \sum_{r=1}^{n-\mathcal{H}_{k+1,j}} B_r^j \right\} - \sum_{i=1}^{K} \sum_{r=1}^{n-\mathcal{N}_{2,k}-1-\mathcal{H}_{2,i}} B_r^i$$

$$
= \max\left\{ \max_{j=1,\dots,k}\left\{ \sum_{r=1}^{n-\mathcal{N}_{j+1,k}} B_r^j - \sum_{i=1}^{K} \sum_{r=1}^{n-\mathcal{N}_{2,k}-1-\mathcal{H}_{2,i}} B_r^i \right\}, \right.
$$
$$
\left. \max_{j=k+1,\dots,K}\left\{ \sum_{r=1}^{n-\mathcal{H}_{k+1,j}} B_r^j - \sum_{i=1}^{K} \sum_{r=1}^{n-\mathcal{N}_{2,k}-1-\mathcal{H}_{2,i}} B_r^i \right\} \right\}
$$

$$
= \max\left\{ \max_{j=1,\dots,k}\left\{ \sum_{r=n-\mathcal{N}_{2,k}-\mathcal{H}_{2,j}}^{n-\mathcal{N}_{j+1,k}} B_r^j - \sum_{\substack{i=1 \\ i\neq j}}^{K} \sum_{r=1}^{n-\mathcal{N}_{2,k}-1-\mathcal{H}_{2,i}} B_r^i \right\}, \right.
$$
$$
\left. \max_{j=k+1,\dots,K}\left\{ \sum_{r=n-\mathcal{N}_{2,k}-\mathcal{H}_{2,j}}^{n-\mathcal{H}_{k+1,j}} B_r^j - \sum_{\substack{i=1 \\ i\neq j}}^{K} \sum_{r=1}^{n-\mathcal{N}_{2,k}-1-\mathcal{H}_{2,i}} B_r^i \right\} \right\}
$$

$$
\geq \max\left\{ \max_{j=1,\dots,k}\left\{ \sum_{r=n-\mathcal{N}_{2,k}-\mathcal{H}_{2,j}}^{n-\mathcal{N}_{j+1,k}} B_r^j - \sum_{i=1}^{K} \sum_{r=1}^{n-\mathcal{N}_{2,k}-1-\mathcal{H}_{2,i}} B_r^i \right\}, \right.
$$
$$
\left. \max_{j=k+1,\dots,K}\left\{ \sum_{r=n-\mathcal{N}_{2,k}-\mathcal{H}_{2,j}}^{n-\mathcal{H}_{k+1,j}} B_r^j - \sum_{i=1}^{K} \sum_{r=1}^{n-\mathcal{N}_{2,k}-1-\mathcal{H}_{2,i}} B_r^i \right\} \right\}
$$

$$
= \max\left\{ \max_{j=1,\dots,k} \sum_{r=n-\mathcal{N}_{2,k}-\mathcal{H}_{2,j}}^{n-\mathcal{N}_{j+1,k}} B_r^j, \; \max_{j=k+1,\dots,K} \sum_{r=n-\mathcal{N}_{2,k}-\mathcal{H}_{2,j}}^{n-\mathcal{H}_{k+1,j}} B_r^j \right\} - \sum_{i=1}^{K} \sum_{r=1}^{n-\mathcal{N}_{2,k}-1-\mathcal{H}_{2,i}} B_r^i
$$

Therefore, from Lemmas 3.0.3 and 3.0.4, for all $n \geq \mathcal{M}_{1,k} + \mathcal{H}_{k+1,K}$,

$$
\liminf_{x\to\infty} \frac{\mathbb{P}(R_n^k > x)}{\overline{F}(x)}
$$
$$
\geq \liminf_{x\to\infty} \frac{\mathbb{P}(\sum_{j=1}^{k}\sum_{r=n-\mathcal{N}_{2,k}-\mathcal{H}_{2,j}}^{n-\mathcal{N}_{j+1,k}} B_r^j + \sum_{j=k+1}^{K}\sum_{r=n-\mathcal{N}_{2,k}-\mathcal{H}_{2,j}}^{n-\mathcal{H}_{k+1,j}} B_r^j > x)}{\overline{F}(x)}
$$
$$
= \liminf_{x\to\infty} \frac{\mathbb{P}(\sum_{j=1}^{k}\sum_{r=1}^{\mathcal{M}_{2,j}+1} B_r^j + \sum_{j=k+1}^{K}\sum_{r=1}^{\mathcal{M}_{2,k}+1} B_r^j > x)}{\overline{F}(x)}
$$
$$
= \sum_{j=1}^{k} \mathcal{M}_{1,j}\, c_j + \mathcal{M}_{1,k} \sum_{j=k+1}^{K} c_j
$$

This together with (30) completes the proof. □

Note that R_n^K is the sojourn time of the nth customer. Then, we have the following simple expression for tail asymptotics of sojourn times. For all $n \geq \mathcal{M}_{1,K}$,

$$
\lim_{x\to\infty} \frac{\mathbb{P}(R_n^K > x)}{\overline{F}(x)} = \sum_{j=1}^{K} \mathcal{M}_{1,j}\, c_j
$$

where the convergence is uniform in n.

Unlike Chapter 4 and 5, let W_n^k denote the time that the nth customer spends at station k, for $k = 1, 2, \ldots, K$. Thus, W_n^k is the time from the arrival of the nth customer at the kth station till its departure from the kth station. Then,

$$W_n^k = X_n^k - X_{n-N_k}^{k-1} \tag{31}$$

for all $k = 2, \ldots, K$.

Since there are infinite number of customers in front of station 1, we have

$$W_n^1 = X_n^1 - X_{n-1}^1. \tag{32}$$

The next proposition provides the tail asymptotics of the nth waiting time at station $k \in \{1, \ldots, K\}$.

Proposition 6.2.5 *For all $k \in \{1, \ldots, K\}$ and all $n \geq M_k + \mathcal{H}_{k+1,K}$, if $B_j \in \mathcal{S}$ for some $j \in \{k, \ldots, K\}$,*

$$\lim_{x \to \infty} \frac{\mathbb{P}(W_n^k > x)}{\overline{F}(x)} = M_k \sum_{j=k}^{K} c_j$$

where the convergence is uniform in n.

Proof We first obtain an upper bound on W_n^k. Clearly, W_n^1 will attain its largest value if all the stations (except station K) are blocked at the time that n^{th} service starts at station 1. Similarly, for $k \in \{2, \ldots, K\}$, if there are $M_k - 1$ customers waiting in front of station k at the time that the nth customer joins station k (i.e., the $(n - M_k + 1)^{th}$ customer is still in service at station k) and the $(n - M_k + 1)^{th}$ customer and all the customers behind him (at station k) get blocked (which will only happen if $k \neq K$), W_n^k will attain its largest value. This argument immediately gives the following upper bound on W_n^k for $k \in \{1, \ldots, K\}$:

$$
\begin{aligned}
W_n^k &\leq \sum_{r=n-M_k+1}^{n} \{B_r^k + \sum_{j=k+1}^{K} B_{r-\mathcal{H}_{k+1,j}}^j\} \\
&= \sum_{r=n-M_k+1}^{n} B_r^k + \sum_{j=k+1}^{K} \sum_{i=n-M_k-\mathcal{H}_{k+1,j}+1}^{n-\mathcal{H}_{k+1,j}} B_i^j.
\end{aligned}
$$

74

Hence, from Corollary 3.0.1, for all $n \geq M_k + \mathcal{H}_{k+1,K}$,

$$
\begin{aligned}
\limsup_{x\to\infty} \frac{\mathbb{P}(W_n^k > x)}{\overline{F}(x)} &\leq \limsup_{x\to\infty} \frac{\mathbb{P}(\sum_{r=n-M_k+1}^{n} B_r^k + \sum_{j=k+1}^{K} \sum_{i=n-M_k-\mathcal{H}_{k+1,j}+1}^{n-\mathcal{H}_{k+1,j}} B_i^j > x)}{\overline{F}(x)} \\
&= \limsup_{x\to\infty} \frac{\mathbb{P}(\sum_{j=k}^{K} \sum_{i=1}^{M_k} B_i^j > x)}{\overline{F}(x)} \\
&= M_k \sum_{j=k}^{K} c_j. \qquad\qquad (33)
\end{aligned}
$$

We now provide a lower bound on W_n^k for $k \in \{1, \ldots, K\}$. From Propositions 6.2.2 and 6.2.3 and equations (31) and (32), we have

$$
\begin{aligned}
W_n^k &\geq \max\Big\{ \max_{j=1,\ldots,k} \sum_{r=1}^{n-\mathcal{N}_{j+1,k}} B_r^j, \max_{j=k+1,\ldots,K} \sum_{r=1}^{n-\mathcal{H}_{k+1,j}} B_r^j \Big\} \\
&\quad - \Big(\sum_{i=1}^{k-1} \sum_{r=1}^{n-N_k-\mathcal{N}_{i+1,k-1}} B_r^i + \sum_{i=k}^{K} \sum_{r=1}^{n-N_k-\mathcal{H}_{k,i}} B_r^i \Big) \\
&= \max\Big\{ \max_{j=1,\ldots,k} \sum_{r=1}^{n-\mathcal{N}_{j+1,k}} B_r^j, \max_{j=k+1,\ldots,K} \sum_{r=1}^{n-\mathcal{H}_{k+1,j}} B_r^j \Big\} \\
&\quad - \Big(\sum_{i=1}^{k-1} \sum_{r=1}^{n-\mathcal{N}_{i+1,k}} B_r^i + \sum_{i=k}^{K} \sum_{r=1}^{n-N_k-\mathcal{H}_{k,i}} B_r^i \Big) \\
&= \max\Big\{ \max_{j=1,\ldots,k-1} \sum_{r=1}^{n-\mathcal{N}_{j+1,k}} B_r^j, \max_{j=k,\ldots,K} \sum_{r=1}^{n-\mathcal{H}_{k+1,j}} B_r^j \Big\} \\
&\quad - \Big(\sum_{i=1}^{k-1} \sum_{r=1}^{n-\mathcal{N}_{i+1,k}} B_r^i + \sum_{i=k}^{K} \sum_{r=1}^{n-N_k-\mathcal{H}_{k,i}} B_r^i \Big) \\
&= \max\Big\{ \max_{j=1,\ldots,k-1}\{ \sum_{r=1}^{n-\mathcal{N}_{j+1,k}} B_r^j - (\sum_{i=1}^{k-1} \sum_{r=1}^{n-\mathcal{N}_{i+1,k}} B_r^i + \sum_{i=k}^{K} \sum_{r=1}^{n-N_k-\mathcal{H}_{k,i}} B_r^i) \}, \\
&\qquad \max_{j=k,\ldots,K}\{ \sum_{r=1}^{n-\mathcal{H}_{k+1,j}} B_r^j - (\sum_{i=1}^{k-1} \sum_{r=1}^{n-\mathcal{N}_{i+1,k}} B_r^i + \sum_{i=k}^{K} \sum_{r=1}^{n-N_k-\mathcal{H}_{k,i}} B_r^i) \} \Big\} \\
&\geq \max\Big\{ \max_{j=1,\ldots,k-1}\{ -(\sum_{\substack{i=1 \\ i\neq j}}^{k-1} \sum_{r=1}^{n-\mathcal{N}_{i+1,k}} B_r^i + \sum_{i=k}^{K} \sum_{r=1}^{n-N_k-\mathcal{H}_{k,i}} B_r^i) \}, \\
&\qquad \max_{j=k,\ldots,K}\{ \sum_{r=n-N_k-\mathcal{H}_{k,j}+1}^{n-\mathcal{H}_{k+1,j}} B_r^j - (\sum_{i=1}^{k-1} \sum_{r=1}^{n-\mathcal{N}_{i+1,k}} B_r^i + \sum_{\substack{i=k \\ i\neq j}}^{K} \sum_{r=1}^{n-N_k-\mathcal{H}_{k,i}} B_r^i) \} \Big\}
\end{aligned}
$$

75

$$\geq \max\left\{ \max_{j=1,\ldots,k-1}\{-(\sum_{i=1}^{k-1}\sum_{r=1}^{n-\mathcal{N}_{i+1,k}}B_r^i + \sum_{i=k}^{K}\sum_{r=1}^{n-N_k-\mathcal{H}_{k,i}}B_r^i)\},\right.$$

$$\left.\max_{j=k,\ldots,K}\{\sum_{r=n-N_k-\mathcal{H}_{k,j}+1}^{n-\mathcal{H}_{k+1,j}}B_r^j - (\sum_{i=1}^{k-1}\sum_{r=1}^{n-\mathcal{N}_{i+1,k}}B_r^i + \sum_{i=k}^{K}\sum_{r=1}^{n-N_k-\mathcal{H}_{k,i}}B_r^i)\}\right\}$$

$$= \max_{j=k,\ldots,K}\{\sum_{r=n-N_k-\mathcal{H}_{k,j}+1}^{n-\mathcal{H}_{k+1,j}}B_r^j - (\sum_{i=1}^{k-1}\sum_{r=1}^{n-\mathcal{N}_{i+1,k}}B_r^i + \sum_{i=k}^{K}\sum_{r=1}^{n-N_k-\mathcal{H}_{k,i}}B_r^i)\}.$$

Therefore, from Lemmas 3.0.3 and 3.0.4, for all $n \geq M_k + \mathcal{H}_{k+1,K}$,

$$\liminf_{x\to\infty}\frac{\mathbb{P}(W_n^k > x)}{\overline{F}(x)}$$

$$\geq \liminf_{x\to\infty}\mathbb{P}\left(\max_{j=k,\ldots,K}\{\sum_{r=n-N_k-\mathcal{H}_{k,j}+1}^{n-\mathcal{H}_{k+1,j}}B_r^j - (\sum_{i=1}^{k-1}\sum_{r=1}^{n-\mathcal{N}_{i+1,k}}B_r^i + \sum_{i=k}^{K}\sum_{r=1}^{n-N_k-\mathcal{H}_{k,i}}B_r^i)\} > x\right)\Big/\overline{F}(x)$$

$$= \liminf_{x\to\infty}\frac{\mathbb{P}(\sum_{j=k}^{K}\sum_{r=n-N_k-\mathcal{H}_{k,j}+1}^{n-\mathcal{H}_{k+1,j}}B_r^j > x)}{\overline{F}(x)}$$

$$= \liminf_{x\to\infty}\frac{\mathbb{P}(\sum_{j=k}^{K}\sum_{i=1}^{M_k}B_i^j > x)}{\overline{F}(x)}$$

$$= M_k\sum_{j=k}^{K}c_j$$

which together with (33) completes the proof. $\qquad\square$

Note that the tail asymptotics of W_n^k only depends on the service times at stations k to K.

Remark 6.2.1 One can easily see that $R_n^k = \sum_{i=1}^{k}W_n^i$ for all $k \in \{1,\ldots,K\}$. However, we would like to point out that the following counter intuitive equality also holds

$$\sum_{i=1}^{k}\lim_{x\to\infty}\frac{\mathbb{P}(W_n^i > x)}{\overline{F}(x)} = \sum_{i=1}^{k}M_i\sum_{j=i}^{K}c_j$$

$$= \sum_{j=1}^{k}\mathcal{M}_{1,j}\,c_j + \mathcal{M}_{1,k}\sum_{j=k+1}^{K}c_j$$

$$= \lim_{x\to\infty}\frac{\mathbb{P}(R_n^k > x)}{\overline{F}(x)}.$$

6.3 Manufacturing Blocking

In this section, we study the tandem queue discussed in Section 6.1 when it operates under the manufacturing blocking scheme. Under this control strategy, at the completion of a service at station k, the customer moves into the buffer of station $k+1$, if that buffer is not full. Otherwise, it has to wait with server k until the downstream buffer has a free space. Hence, unlike communication blocking a customer gets blocked after service.

As is done in Section 6.2, we first obtain a recursive relationship for the departure time of the nth customer from station k, namely X_n^k for all $k \in \{1, \ldots, K\}$.

Proposition 6.3.1 *For all $k \in \{1, \ldots, K\}$ and $n \geq 1$,*

$$X_n^k = \max\{X_{n-1}^k + B_n^k, \; X_{n-N_k}^{k-1} + B_n^k, \; X_{n-H_{k+1}}^{k+1}\} \tag{34}$$

with the convention that $X_n^0 = 0$, $X_n^{K+1} = 0$ for all n and $X_n^k = 0$ for all $n \leq 0$ and all $k \in \{1, \ldots, K\}$.

Proof At the time of the nth service completion at station $k \in \{1, \ldots, K\}$, if buffer $k+1$ is not full, the nth departure time is the maximum of $X_{n-1}^k + B_n^k$ and $X_{n-N_k}^{k-1} + B_n^k$. If buffer $k+1$ is full, the nth customer is blocked and needs to wait for the blocking to be cleared. $\square$

One can obtain the following upper bound on X_n^k by summing up all the service times that appear in (34).

Proposition 6.3.2 *For all $k \in \{1, \ldots, K\}$ and $n \geq 1$,*

$$X_n^k \leq \sum_{u=1}^{k} \sum_{r=1}^{n-\mathcal{N}_{u+1,k}} B_r^u + \sum_{u=k+1}^{K} \sum_{r=1}^{n-\mathcal{H}_{k+1,u}} B_r^u$$

with the convention that the summation over an empty set is zero.

Similarly, employing the techniques used in the proof of Proposition 6.2.3, we obtain the following lower bound on departure times.

Proposition 6.3.3 *For all $k \in \{1, \ldots, K\}$ and $n \geq 1$,*

$$X_n^k \geq \max\{\max_{u=1,\ldots,k} \sum_{r=1}^{n-\mathcal{N}_{u+1,k}} B_r^u, \max_{u=k+1,\ldots,K} \sum_{r=1}^{n-\mathcal{H}_{k+1,u}} B_r^u\}$$

with the convention that the summation over an empty set is equal to zero and the maximization over an empty set is equal to $-\infty$.

Note that X_n^k can be bounded above and below by the same expressions under both blocking strategies which is not surprising since the recursive expression for X_n^k under both blocking schemes is similar. Propositions 6.3.4 and 6.3.5 provide the tail asymptotics of transient response times and waiting times, respectively for tandem lines with manufacturing blocking. As the results illustrate tail asymptotics for both performance measures are the same under both blocking schemes.

Proposition 6.3.4 *For all $k \in \{1, \ldots, K\}$ and all $n \geq \mathcal{M}_{1,k} + \mathcal{H}_{k+1,K}$,*

$$\lim_{x \to \infty} \frac{\mathbb{P}(R_n^k > x)}{\overline{F}(x)} = \sum_{j=1}^{k} \mathcal{M}_{1,j}\, c_j + \mathcal{M}_{1,k} \sum_{j=k+1}^{K} c_j$$

where the convergence is uniform in n.

Proof Note that R_n^k will attain its largest value if all the stations (except station K) are blocked at the time that the $(n-\mathcal{N}_{2,k})^{th}$ customer departs from station 1. That is the $(n - \mathcal{N}_{2,k} - \mathcal{H}_{2,K})^{th}$ customer is still at station K when the $(n - \mathcal{N}_{2,k})^{th}$ customer departs from station 1. Since $\mathcal{N}_{2,k} + \mathcal{H}_{2,j} = \mathcal{M}_{2,k} + \mathcal{H}_{k+1,j}$ for all $j \in \{k+1, \ldots, K\}$, we immediately obtain the following upper bound on R_n^k.

$$R_n^k \leq \sum_{j=1}^{k} \sum_{r=n-\mathcal{N}_{j+1,k}-\mathcal{M}_{2,j}}^{n-\mathcal{N}_{j+1,k}} B_r^j + \sum_{j=k+1}^{K} \sum_{r=n-\mathcal{M}_{2,k}-\mathcal{H}_{k+1,j}}^{n-\mathcal{H}_{k+1,j}} B_r^j.$$

Hence, from Corollary 3.0.1, for all $n \geq \mathcal{M}_{1,k} + \mathcal{H}_{k+1,K}$,

$$\limsup_{x \to \infty} \frac{\mathbb{P}(R_n^k > x)}{\overline{F}(x)}$$

$$\leq \limsup_{x \to \infty} \frac{\mathbb{P}(\sum_{j=1}^{k} \sum_{r=n-\mathcal{N}_{j+1,k}-\mathcal{M}_{2,j}}^{n-\mathcal{N}_{j+1,k}} B_r^j + \sum_{j=k+1}^{K} \sum_{r=n-\mathcal{M}_{2,k}-\mathcal{H}_{k+1,j}}^{n-\mathcal{H}_{k+1,j}} B_r^j > x)}{\overline{F}(x)}$$

78

$$
= \limsup_{x \to \infty} \frac{\mathbb{P}(\sum_{j=1}^{k} \sum_{r=1}^{\mathcal{M}_{2,j}+1} B_r^j + \sum_{j=k+1}^{K} \sum_{r=1}^{\mathcal{M}_{2,k}+1} B_r^j > x)}{\overline{F}(x)}
$$

$$
= \sum_{j=1}^{k} \mathcal{M}_{1,j}\, c_j + \mathcal{M}_{1,k} \sum_{j=k+1}^{K} c_j. \tag{35}
$$

We now provide a lower bound on R_n^k. From Propositions 6.3.2 and 6.3.3, we have

$$
X_n^k - X_{n-\mathcal{N}_{2,k}-1}^1
$$

$$
\geq \max \left\{ \max_{j=1,\ldots,k} \sum_{r=1}^{n-\mathcal{N}_{j+1,k}} B_r^j, \ \max_{j=k+1,\ldots,K} \sum_{r=1}^{n-\mathcal{H}_{k+1,j}} B_r^j \right\} - \sum_{i=1}^{K} \sum_{r=1}^{n-\mathcal{N}_{2,k}-1-\mathcal{H}_{2,i}} B_r^i
$$

$$
= \max \left\{ \max_{j=1,\ldots,k} \left\{ \sum_{r=1}^{n-\mathcal{N}_{j+1,k}} B_r^j - \sum_{i=1}^{K} \sum_{r=1}^{n-\mathcal{N}_{2,k}-1-\mathcal{H}_{2,i}} B_r^i \right\}, \right.
$$

$$
\left. \max_{j=k+1,\ldots,K} \left\{ \sum_{r=1}^{n-\mathcal{H}_{k+1,j}} B_r^j - \sum_{i=1}^{K} \sum_{r=1}^{n-\mathcal{N}_{2,k}-1-\mathcal{H}_{2,i}} B_r^i \right\} \right\}
$$

$$
= \max \left\{ \max_{j=1,\ldots,k} \left\{ \sum_{r=n-\mathcal{N}_{2,k}-\mathcal{H}_{2,j}}^{n-\mathcal{N}_{j+1,k}} B_r^j - \sum_{\substack{i=1 \\ i \neq j}}^{K} \sum_{r=1}^{n-\mathcal{N}_{2,k}-1-\mathcal{H}_{2,i}} B_r^i \right\}, \right.
$$

$$
\left. \max_{j=k+1,\ldots,K} \left\{ \sum_{r=n-\mathcal{N}_{2,k}-\mathcal{H}_{2,j}}^{n-\mathcal{H}_{k+1,j}} B_r^j - \sum_{\substack{i=1 \\ i \neq j}}^{K} \sum_{r=1}^{n-\mathcal{N}_{2,k}-1-\mathcal{H}_{2,i}} B_r^i \right\} \right\}
$$

$$
\geq \max \left\{ \max_{j=1,\ldots,k} \left\{ \sum_{r=n-\mathcal{N}_{2,k}-\mathcal{H}_{2,j}}^{n-\mathcal{N}_{j+1,k}} B_r^j - \sum_{i=1}^{K} \sum_{r=1}^{n-\mathcal{N}_{2,k}-1-\mathcal{H}_{2,i}} B_r^i \right\}, \right.
$$

$$
\left. \max_{j=k+1,\ldots,K} \left\{ \sum_{r=n-\mathcal{N}_{2,k}-\mathcal{H}_{2,j}}^{n-\mathcal{H}_{k+1,j}} B_r^j - \sum_{i=1}^{K} \sum_{r=1}^{n-\mathcal{N}_{2,k}-1-\mathcal{H}_{2,i}} B_r^i \right\} \right\}
$$

$$
= \max \left\{ \max_{j=1,\ldots,k} \sum_{r=n-\mathcal{N}_{2,k}-\mathcal{H}_{2,j}}^{n-\mathcal{N}_{j+1,k}} B_r^j, \ \max_{j=k+1,\ldots,K} \sum_{r=n-\mathcal{N}_{2,k}-\mathcal{H}_{2,j}}^{n-\mathcal{H}_{k+1,j}} B_r^j \right\} - \sum_{i=1}^{K} \sum_{r=1}^{n-\mathcal{N}_{2,k}-1-\mathcal{H}_{2,i}} B_r^i
$$

Therefore, from Lemmas 3.0.3 and 3.0.4, for all $n \geq \mathcal{M}_{1,k} + \mathcal{H}_{k+1,K}$,

$$
\liminf_{x \to \infty} \frac{\mathbb{P}(R_n^k > x)}{\overline{F}(x)}
$$

$$
\geq \liminf_{x \to \infty} \frac{\mathbb{P}(\sum_{j=1}^{k} \sum_{r=n-\mathcal{N}_{2,k}-\mathcal{H}_{2,j}}^{n-\mathcal{N}_{j+1,k}} B_r^j + \sum_{j=k+1}^{K} \sum_{r=n-\mathcal{N}_{2,k}-\mathcal{H}_{2,j}}^{n-\mathcal{H}_{k+1,j}} B_r^j > x)}{\overline{F}(x)}
$$

$$= \liminf_{x \to \infty} \frac{\mathbb{P}(\sum_{j=1}^{k} \sum_{r=1}^{\mathcal{M}_{2,j}+1} B_r^j + \sum_{j=k+1}^{K} \sum_{r=1}^{\mathcal{M}_{2,k}+1} B_r^j > x)}{\overline{F}(x)}$$

$$= \sum_{j=1}^{k} \mathcal{M}_{1,j} \, c_j + \mathcal{M}_{1,k} \sum_{j=k+1}^{K} c_j$$

This together with (35) completes the proof. □

Proposition 6.3.5 *For all $k \in \{1, \ldots, K\}$ and all $n \geq M_k + \mathcal{H}_{k+1,K}$, if $B_j \in \mathcal{S}$ for some $j \in \{k, \ldots, K\}$*

$$\lim_{x \to x} \frac{\mathbb{P}(W_n^k > x)}{\overline{F}(x)} = M_k \sum_{j=k}^{K} c_j$$

where the convergence is uniform in n.

Proof We first obtain an upper bound on W_n^k. Clearly, W_n^1 will attain its largest value if all the stations (except station K) are blocked at the time that the nth customer departs from station 1. Similarly, for $k \in \{2, \ldots, K\}$, if there are $M_k - 1$ customers waiting in front of station k at the time that the nth customer joins station k (i.e., the $(n - M_k + 1)^{th}$ customer is still in service at station k) and the $(n - M_k + 1)^{th}$ customer and all the customers behind him (at station k) get blocked (which will only happen if $k \neq K$), W_n^k will attain its largest value. This argument immediately gives the following upper bound on W_n^k for $k \in \{1, \ldots, K\}$:

$$W_n^k \leq \sum_{r=n-M_k+1}^{n} \{B_r^k + \sum_{j=k+1}^{K} B_{r-\mathcal{H}_{k+1,j}}^j\}$$

$$= \sum_{r=n-M_k+1}^{n} B_r^k + \sum_{j=k+1}^{K} \sum_{i=n-M_k-\mathcal{H}_{k+1,j}+1}^{n-\mathcal{H}_{k+1,j}} B_i^j.$$

Hence, from Corollary 3.0.1, for all $n \geq M_k + \mathcal{H}_{k+1,K}$,

$$\limsup_{x \to \infty} \frac{\mathbb{P}(W_n^k > x)}{\overline{F}(x)} \leq \limsup_{x \to \infty} \frac{\mathbb{P}(\sum_{r=n-M_k+1}^{n} B_r^k + \sum_{j=k+1}^{K} \sum_{i=n-M_k-\mathcal{H}_{k+1,j}+1}^{n-\mathcal{H}_{k+1,j}} B_i^j > x)}{\overline{F}(x)}$$

$$= \limsup_{x \to \infty} \frac{\mathbb{P}(\sum_{j=k}^{K} \sum_{i=1}^{M_k} B_i^j > x)}{\overline{F}(x)}$$

$$= M_k \sum_{j=k}^{K} c_j. \tag{36}$$

We now provide a lower bound on W_n^k for $k \in \{1, \ldots, K\}$. From Propositions 6.3.2 and 6.3.3 and equations (31) and (32), we have

$$
\begin{aligned}
W_n^k \;\geq\; & \max\left\{ \max_{j=1,\ldots,k} \sum_{r=1}^{n-\mathcal{N}_{j+1,k}} B_r^j, \; \max_{j=k+1,\ldots,K} \sum_{r=1}^{n-\mathcal{H}_{k+1,j}} B_r^j \right\} \\
& - \left(\sum_{i=1}^{k-1} \sum_{r=1}^{n-N_k-\mathcal{N}_{i+1,k-1}} B_r^i + \sum_{i=k}^{K} \sum_{r=1}^{n-N_k-\mathcal{H}_{k,i}} B_r^i \right) \\
=\; & \max\left\{ \max_{j=1,\ldots,k} \sum_{r=1}^{n-\mathcal{N}_{j+1,k}} B_r^j, \; \max_{j=k+1,\ldots,K} \sum_{r=1}^{n-\mathcal{H}_{k+1,j}} B_r^j \right\} \\
& - \left(\sum_{i=1}^{k-1} \sum_{r=1}^{n-\mathcal{N}_{i+1,k}} B_r^i + \sum_{i=k}^{K} \sum_{r=1}^{n-N_k-\mathcal{H}_{k,i}} B_r^i \right) \\
=\; & \max\left\{ \max_{j=1,\ldots,k-1} \sum_{r=1}^{n-\mathcal{N}_{j+1,k}} B_r^j, \; \max_{j=k,\ldots,K} \sum_{r=1}^{n-\mathcal{H}_{k+1,j}} B_r^j \right\} \\
& - \left(\sum_{i=1}^{k-1} \sum_{r=1}^{n-\mathcal{N}_{i+1,k}} B_r^i + \sum_{i=k}^{K} \sum_{r=1}^{n-N_k-\mathcal{H}_{k,i}} B_r^i \right) \\
\geq\; & \max\left\{ \max_{j=1,\ldots,k-1}\left\{ \sum_{r=1}^{n-\mathcal{N}_{j+1,k}} B_r^j - \left(\sum_{i=1}^{k-1} \sum_{r=1}^{n-\mathcal{N}_{i+1,k}} B_r^i + \sum_{i=k}^{K} \sum_{r=1}^{n-N_k-\mathcal{H}_{k,i}} B_r^i \right) \right\}, \right. \\
& \left. \max_{j=k,\ldots,K}\left\{ \sum_{r=1}^{n-\mathcal{H}_{k+1,j}} B_r^j - \left(\sum_{i=1}^{k-1} \sum_{r=1}^{n-\mathcal{N}_{i+1,k}} B_r^i + \sum_{i=k}^{K} \sum_{r=1}^{n-N_k-\mathcal{H}_{k,i}} B_r^i \right) \right\} \right\} \\
\geq\; & \max\left\{ \max_{j=1,\ldots,k-1}\left\{ -\left(\sum_{\substack{i=1 \\ i\neq j}}^{k-1} \sum_{r=1}^{n-\mathcal{N}_{i+1,k}} B_r^i + \sum_{i=k}^{K} \sum_{r=1}^{n-N_k-\mathcal{H}_{k,i}} B_r^i \right) \right\}, \right. \\
& \left. \max_{j=k,\ldots,K}\left\{ \sum_{r=n-N_k-\mathcal{H}_{k,j}+1}^{n-\mathcal{H}_{k+1,j}} B_r^j - \left(\sum_{i=1}^{k-1} \sum_{r=1}^{n-\mathcal{N}_{i+1,k}} B_r^i + \sum_{\substack{i=k \\ i\neq j}}^{K} \sum_{r=1}^{n-N_k-\mathcal{H}_{k,i}} B_r^i \right) \right\} \right\} \\
\geq\; & \max\left\{ \max_{j=1,\ldots,k-1}\left\{ -\left(\sum_{i=1}^{k-1} \sum_{r=1}^{n-\mathcal{N}_{i+1,k}} B_r^i + \sum_{i=k}^{K} \sum_{r=1}^{n-N_k-\mathcal{H}_{k,i}} B_r^i \right) \right\}, \right. \\
& \left. \max_{j=k,\ldots,K}\left\{ \sum_{r=n-N_k-\mathcal{H}_{k,j}+1}^{n-\mathcal{H}_{k+1,j}} B_r^j - \left(\sum_{i=1}^{k-1} \sum_{r=1}^{n-\mathcal{N}_{i+1,k}} B_r^i + \sum_{i=k}^{K} \sum_{r=1}^{n-N_k-\mathcal{H}_{k,i}} B_r^i \right) \right\} \right\} \\
=\; & \max_{j=k,\ldots,K}\left\{ \sum_{r=n-N_k-\mathcal{H}_{k,j}+1}^{n-\mathcal{H}_{k+1,j}} B_r^j - \left(\sum_{i=1}^{k-1} \sum_{r=1}^{n-\mathcal{N}_{i+1,k}} B_r^i + \sum_{i=k}^{K} \sum_{r=1}^{n-N_k-\mathcal{H}_{k,i}} B_r^i \right) \right\}.
\end{aligned}
$$

Therefore, from Lemmas 3.0.3 and 3.0.4, for all $n \geq M_k + \mathcal{H}_{k+1,K}$,

$$
\liminf_{x \to \infty} \frac{\mathbb{P}(W_n^k > x)}{\overline{F}(x)}
$$

$$
\geq \liminf_{x \to \infty} \mathbb{P}\Big(\max_{j=k,\ldots,K} \Big\{ \sum_{r=n-N_k-\mathcal{H}_{k,j}+1}^{n-\mathcal{H}_{k+1,j}} B_r^j - \Big(\sum_{i=1}^{k-1} \sum_{r=1}^{n-\mathcal{N}_{i+1,k}} B_r^i + \sum_{i=k}^{K} \sum_{r=1}^{n-N_k-\mathcal{H}_{k,i}} B_r^i \Big) \Big\} > x \Big) \Big/ \overline{F}(x)
$$

$$
= \liminf_{x \to \infty} \frac{\mathbb{P}(\sum_{j=k}^{K} \sum_{r=n-N_k-\mathcal{H}_{k,j}+1}^{n-\mathcal{H}_{k+1,j}} B_r^j > x)}{\overline{F}(x)}
$$

$$
= \liminf_{x \to \infty} \frac{\mathbb{P}(\sum_{j=k}^{K} \sum_{i=1}^{M_k} B_i^j > x)}{\overline{F}(x)}
$$

$$
= M_k \sum_{j=k}^{K} c_j
$$

which together with (36) completes the proof. $\qquad\square$

6.4   Stationary Response times and Waiting times

In this section, we provide the tail asymptotics of stationary response times and waiting times under both blocking schemes. Let R^k and W^k denote the stationary response time and waiting time at station $k \in \{1, \ldots, K\}$, respectively when they exist. Since the convergence in Propositions 6.2.4, 6.2.5, 6.3.4, and 6.3.5 is uniform in n, we immediately have the following result.

Proposition 6.4.1 *If a stationary regime exists, then under communication blocking and manufacturing blocking, for all $k \in \{1, \ldots, K\}$,*

$$
\lim_{x \to \infty} \frac{\mathbb{P}(R^k > x)}{\overline{F}(x)} = \sum_{j=1}^{k} \mathcal{M}_{1,j} \, c_j + \mathcal{M}_{1,k} \sum_{j=k+1}^{K} c_j,
$$

and for all $k \in \{1, \ldots, K\}$ with $B_j \in \mathcal{S}$ for some $j \in \{k, \ldots, K\}$,

$$
\lim_{x \to \infty} \frac{\mathbb{P}(W^k > x)}{\overline{F}(x)} = M_k \sum_{j=k}^{K} c_j.
$$

We next provide sufficient conditions under which a stationary regime exists. The tandem queue of this chapter under both blocking strategies is an example of a (max,+)

linear system (see Baccelli, Cohen, Olsder and Quadrat [11] for details of (max,+) linear systems). Moreover, even though the network that we study is open, since there are infinite number of customers in front of station 1, we have an autonomous (max,+) linear system (i.e., the evolution equations are the same as the one in equation 7.92 on page 353 of [11]). Then, as in Chapter 4 and 5, using the analysis in Section 7.5 of [11], we can derive sufficient conditions under which the stationary characteristics exist. In particular, Theorem 7.94 of [11] states that if there exists a station $k \in \{1, \ldots, K\}$ that can start processing a customer at time 0 and has a service time distribution with infinite support then the sequence of vectors $\{(X_n^i - X_{n-1}^j) : i, j \in \{1, \ldots, K\}\}_{n \geq 1}$ admits a unique stationary regime which is integrable, directly reachable, independent of the initial condition and $\{(X_n^i - X_{n-1}^j) : i, j \in \{1, \ldots, K\}\}$ couples with it in finite time. Note that both the response times and the waiting times can be expressed in terms of these differences of departure times. In particular, for all $k \in \{1, \ldots, K\}$

$$R_n^k \;=\; X_n^k - X_{n-1}^1 + \sum_{r=n-\mathcal{N}_{2,k}}^{n-1} X_r^1 - X_{r-1}^1,$$

for $k \in \{2, \ldots, K\}$

$$W_n^k = X_n^k - X_{n-1}^{k-1} + \sum_{r=n-N_k+1}^{n-1} X_r^{k-1} - X_{r-1}^{k-1}$$

and by definition $W_n^1 = X_n^k - X_{n-1}^1$.

Then, a set of sufficient conditions for communication blocking is:

(i) $N_2 < M_2$ and $B_1(\cdot)$ has infinite support or

(ii) there exists $k \in \{2, \ldots, K-1\}$ such that $N_k > 0$, $N_{k+1} < M_{k+1}$ and $B_k(\cdot)$ has infinite support or

(iii) $N_K > 0$ and $B_K(\cdot)$ has infinite support.

Similarly, a set of sufficient conditions for manufacturing blocking is:

(i) $B_1(\cdot)$ has infinite support or

(ii) there exists $k \in \{2, \ldots, K\}$ such that $N_k > 0$ and $B_k(\cdot)$ has infinite support.

Note that if the service times at all stations have infinite support (which is clearly satisfied if the service time distributions are subexponential), above sufficient conditions are satisfied and a stationary regime exists.

6.5   Numerical Results

The results in Propositions 6.2.4, 6.2.5, 6.3.4, and 6.3.5 provide the behavior of the tail probabilities for transient response times and waiting times as x gets large. Clearly, one would like to understand how fast the convergence of these tail probabilities is to their asymptotic counter parts. In this section, we provide numerical experiments to study the convergence behavior of the transient response times and waiting times for tandem lines under communication blocking and manufacturing blocking rules.

We consider three systems. System 1 has $K = 5$ stations with $M_k = 5$ for all $k \in \{2, \ldots, 5\}$, and thus, $M = 21$. System 2 has $K = 11$ stations with $M_k = 2$ for all $k \in \{2, \ldots, 11\}$, and thus, $M = 21$. System 3 has $K = 5$ stations with $M_k = 12$ for all $k \in \{2, \ldots, 5\}$, and thus, $M = 49$. Thus, Systems 1 and 2 have the same capacity but different number of stations. On the other hand, Systems 1 and 3 have the same number of stations but different capacities. The number of initial customers in all three systems is zero. We assume that service time distributions at all stations for all three systems are Pareto with parameter 1 (i.e., $\overline{B}_k(x) = x^{-1}$ for all $k \in \{1, \ldots, K\}$). Thus, $c_k = 1$ for all $k \in \{1, \ldots, K\}$. Under these assumptions, the tail asymptotics of transient response times and waiting times are both independent of n as long as $n \geq M$ as discussed in Section 6.2.2. We consider the sojourn time (response time corresponding to the last station) and waiting times (at various stations) of the 50th customer in all three systems. In particular, for each value of x, we first approximate the tail probabilities of sojourn time and waiting time using the tail asymptotics of sojourn time and waiting time as given in Propositions 6.2.4 and 6.2.5, respectively and compare them with the tail probabilities obtained from simulation analysis. In

our simulation study, for each value of x we run 41 batches of 10,000 replications and compute the average and 95% confidence interval of the corresponding tail probability. However, for purposes of clarity, we do not present the confidence intervals in Figures 42 to 57.

We first focus on Systems 1 and 2 with communication blocking, which have the same system capacity but different number of stations . For System 1, we compute the tail probabilities of sojourn time and waiting times at stations 1, 3 (middle station), and 5 (end station) of the 50th customer. Similarly, for System 2 we compute the tail probabilities of sojourn time and waiting times at stations 1, 6 (middle station), and 11 (end station) of the 50th customer. Figures 42 through 45 illustrate how these tail probabilities vary with respect to x. As Figure 42 demonstrates for systems with small number of stations (i.e., System 1) the tail asymptotics could provide a good approximation for the tail probabilities of sojourn times even when x is moderately large. However, the convergence of the tail asymptotics to the actual tail probability is slower for System 2 which is expected since the tail asymptotics of the sojourn time depend on the number of stations. Figures 43 to 45 present the tail probabilities of waiting times of the 50th customer at the first station, the middle station, and the last station for Systems 1 and 2. As the figures demonstrate, in all these cases the tail asymptotics could be used to approximate the tail probabilities of waiting times even when x is moderately large. Note that the convergence of the tail asymptotics to the tail probabilities is especially fast for the waiting times at the last station.

We next focus on Systems 1 and 3 with communication blocking, which have the same number of stations but different system capacities. For both systems we again compute the tail probabilities of sojourn time and waiting times at stations 1, 3 (middle station), and 5 (end station) of the 50th customer. Figures 46 through 49 illustrate how these tail probabilities vary with respect to x. As Figure 46 demonstrates the
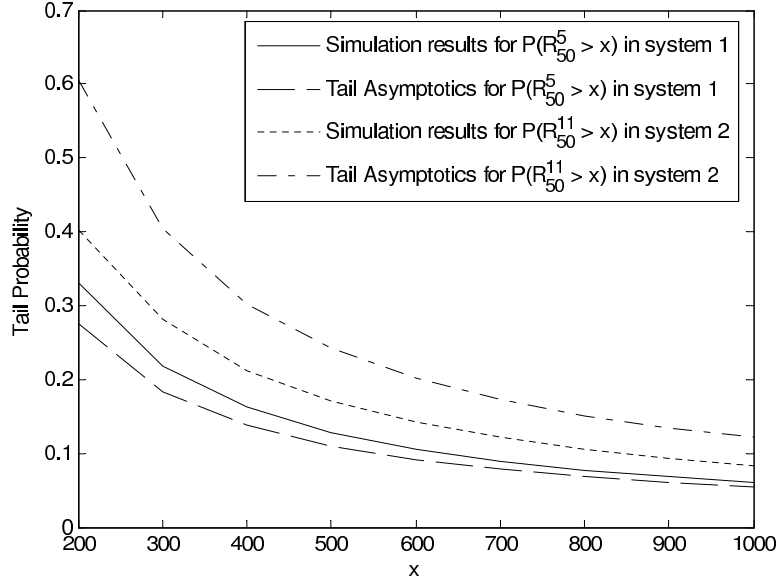
Figure 42: Simulation results and the tail asymptotics of sojourn times for Systems 1 and 2 with communication blocking

tail asymptotics could provide a good approximation for the tail probabilities of sojourn times for both systems even when x is moderately large but the approximation is especially good for the system with smaller capacity (i.e., System 1). Figures 47 to 49 present the tail probabilities of waiting times of the 50th customer at station 1, station 3, and station 5 for Systems 1 and 3. Note that in Figure 47 tail asymptotics of the waiting times at the first station in both systems are the same since both systems have the same number of stations and the capacity of station 1 is 1. As we have observed above, in all these cases the tail asymptotics could provide a good approximation for the tail probability of the waiting times as x increases from medium to large values and the convergence is again especially fast for the waiting times at the last station.

Figures 50 through 57 present the tail asymptotics and simulation results of transient sojourn times and waiting times for all three systems operating manufacturing blocking. These figures show that the numerical results for communication blocking and manufacturing blocking are very similar.
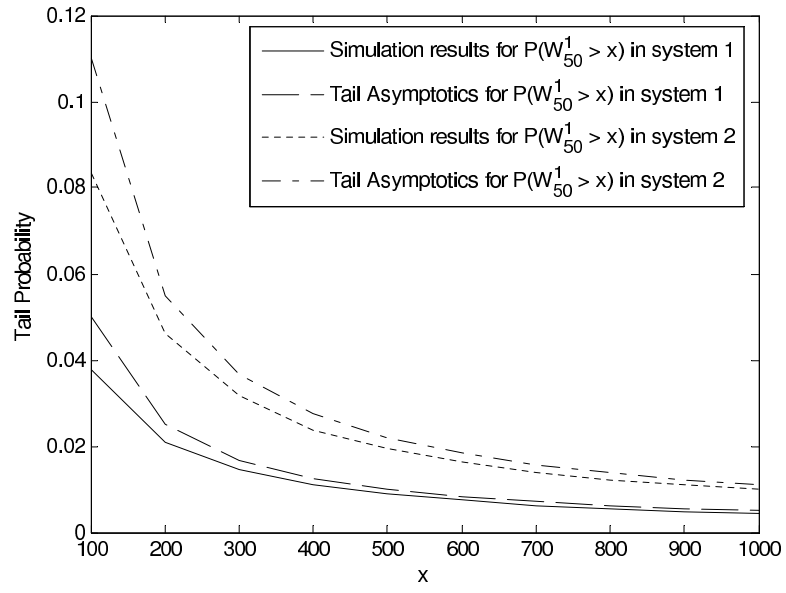
Figure 43: Simulation results and the tail asymptotics of waiting times of the first station for Systems 1 and 2 with communication blocking
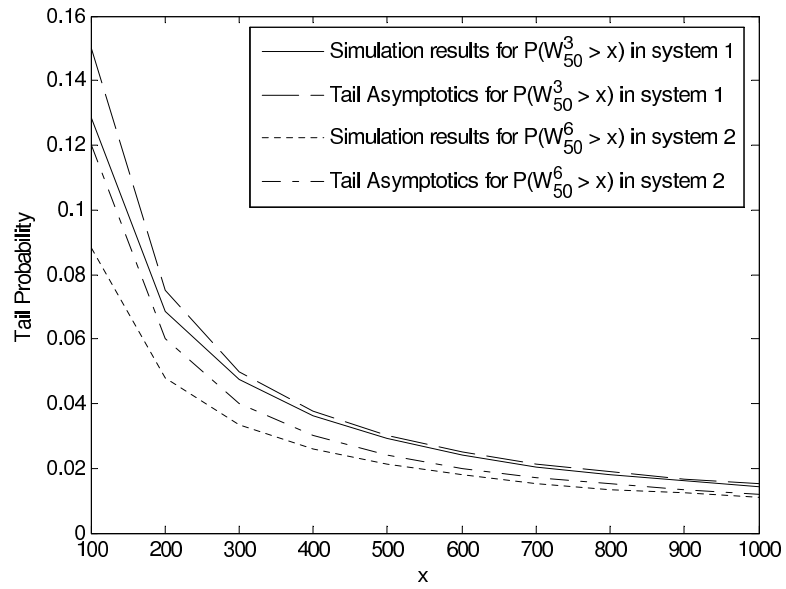

Figure 44: Simulation results and the tail asymptotics of waiting times of the middle station for Systems 1 and 2 with communication blocking
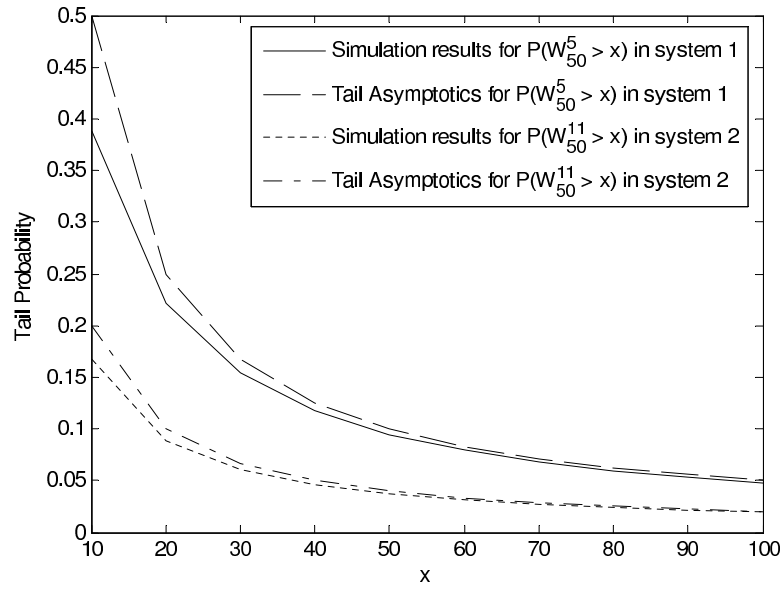
87

Figure 45: Simulation results and the tail asymptotics of waiting times of the last station for Systems 1 and 2 with communication blocking
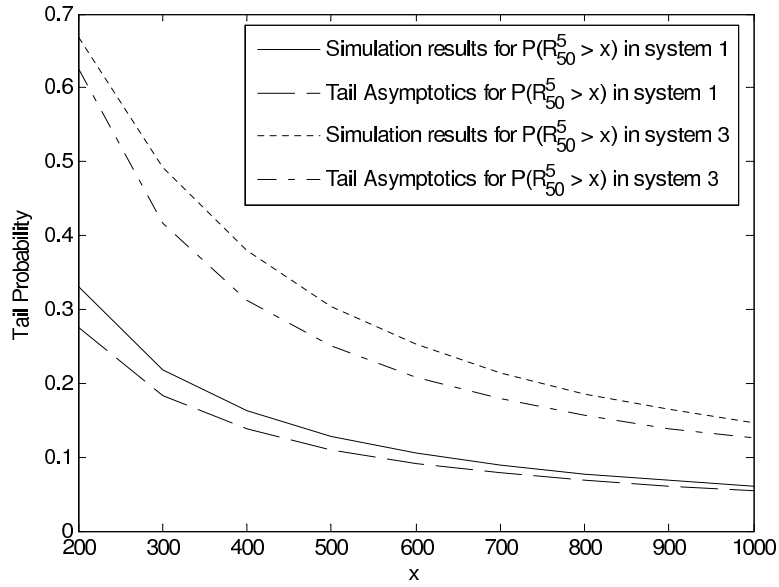

Figure 46: Simulation results and the tail asymptotics of sojourn times for Systems 1 and 3 with communication blocking
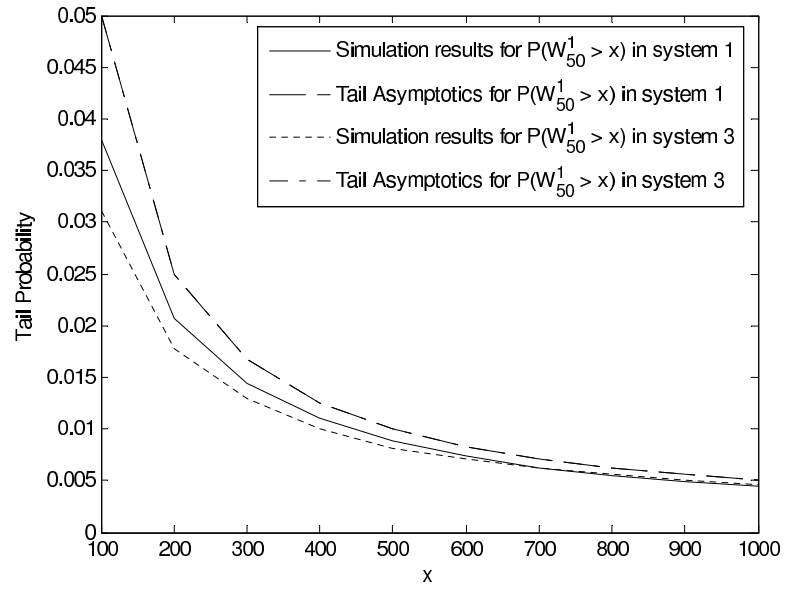
Figure 47: Simulation results and the tail asymptotics of waiting times of the first station for Systems 1 and 3 with communication blocking
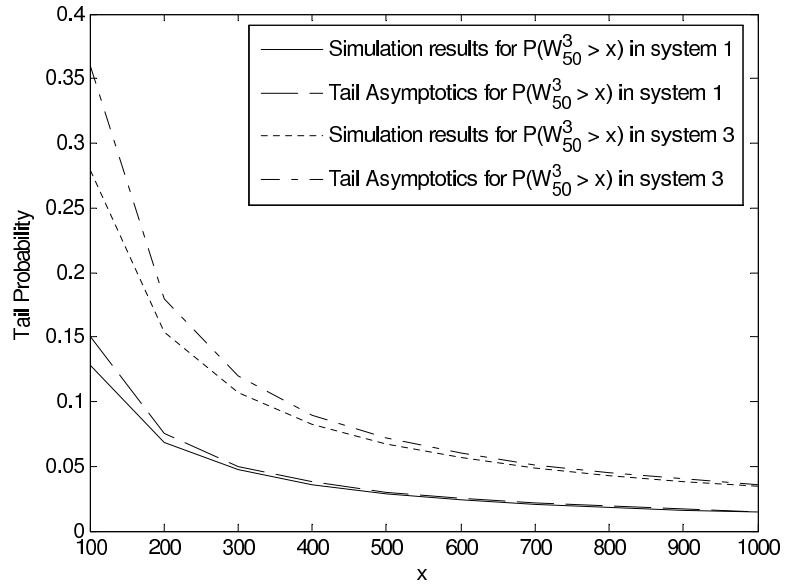

Figure 48: Simulation results and the tail asymptotics of waiting times of the middle station for Systems 1 and 3 with communication blocking
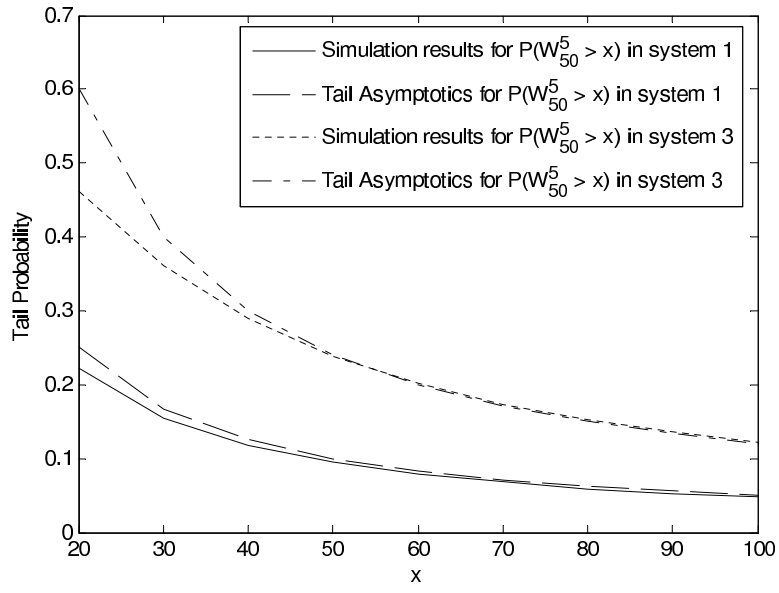
89

Figure 49: Simulation results and the tail asymptotics of waiting times of the last station for Systems 1 and 3 with communication blocking
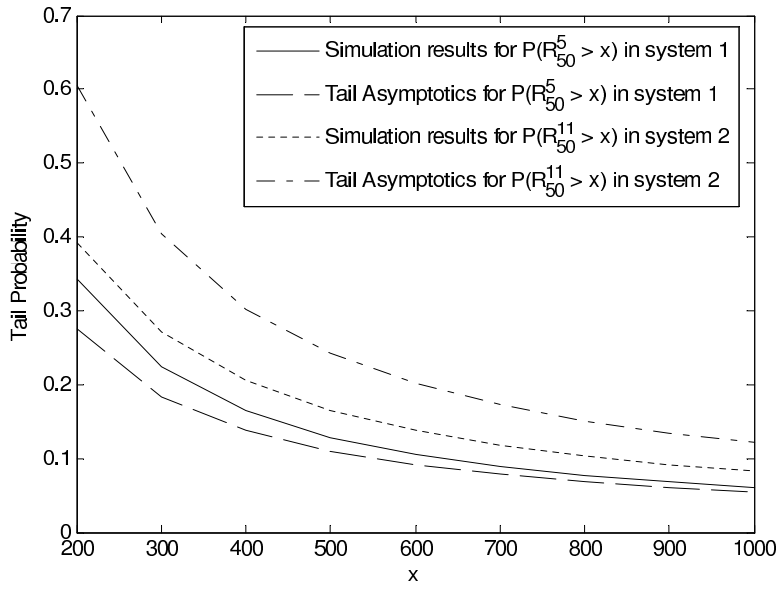

Figure 50: Simulation results and the tail asymptotics of sojourn times for Systems 1 and 2 with manufacturing blocking

Figure 51: Simulation results and the tail asymptotics of waiting times of the first station for Systems 1 and 2 with manufacturing blocking

Figure 52: Simulation results and the tail asymptotics of waiting times of the middle station for Systems 1 and 2 with manufacturing blocking

91

Figure 53: Simulation results and the tail asymptotics of waiting times of the last station for Systems 1 and 2 with manufacturing blocking

Figure 54: Simulation results and the tail asymptotics of sojourn times for Systems 1 and 3 with manufacturing blocking
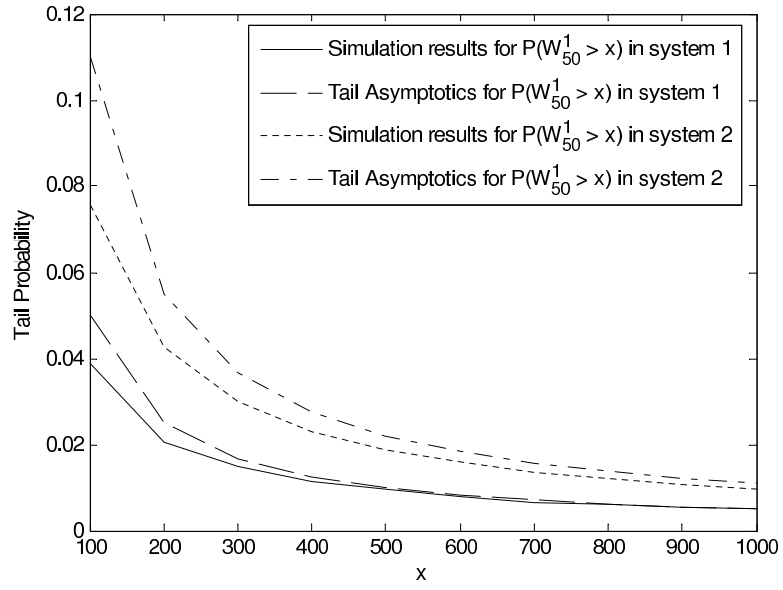
Figure 55: Simulation results and the tail asymptotics of waiting times of the first station for Systems 1 and 3 with manufacturing blocking
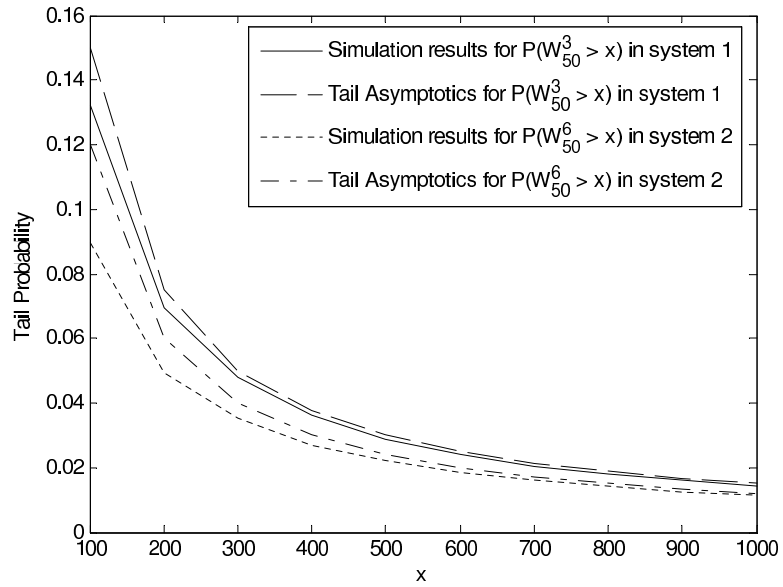

Figure 56: Simulation results and the tail asymptotics of waiting times of the middle station for Systems 1 and 3 with manufacturing blocking
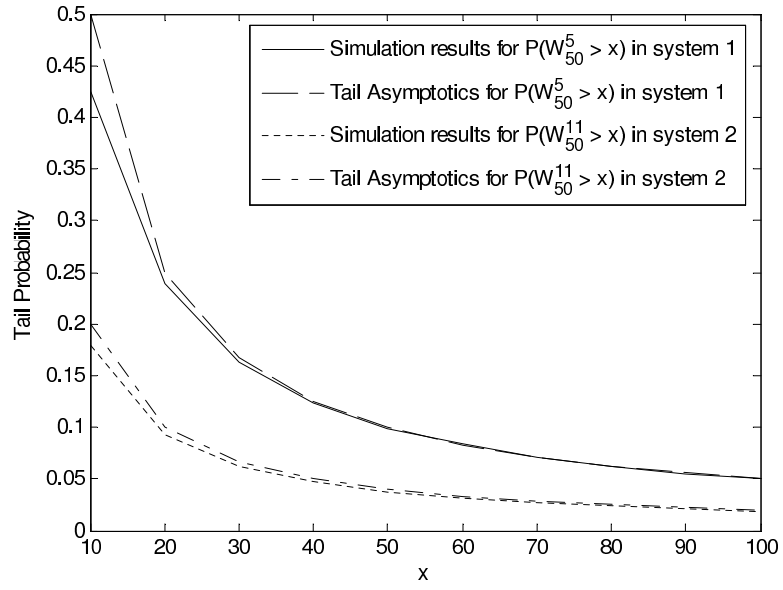
93

Figure 57: Simulation results and the tail asymptotics of waiting times of the last station for Systems 1 and 3 with manufacturing blocking
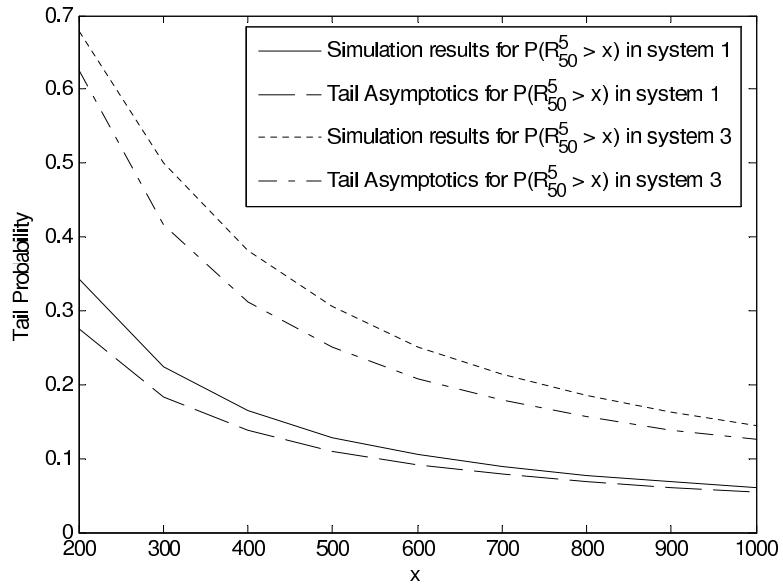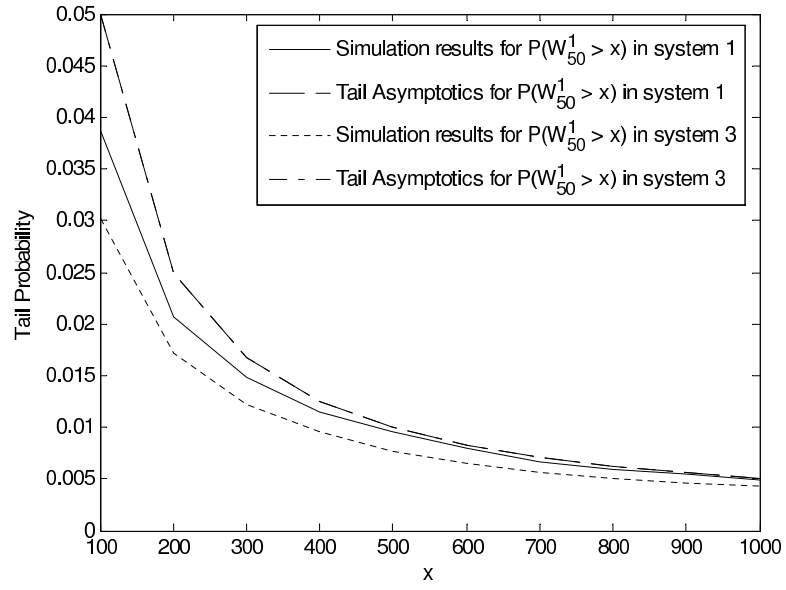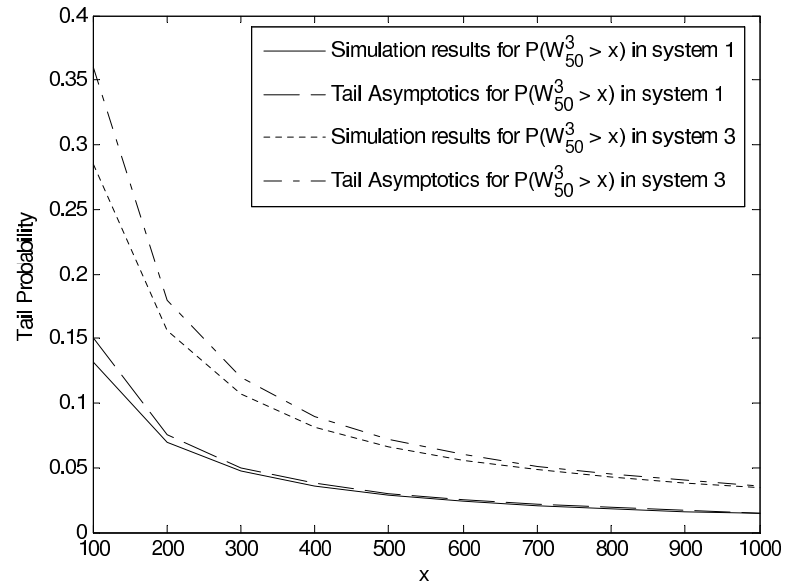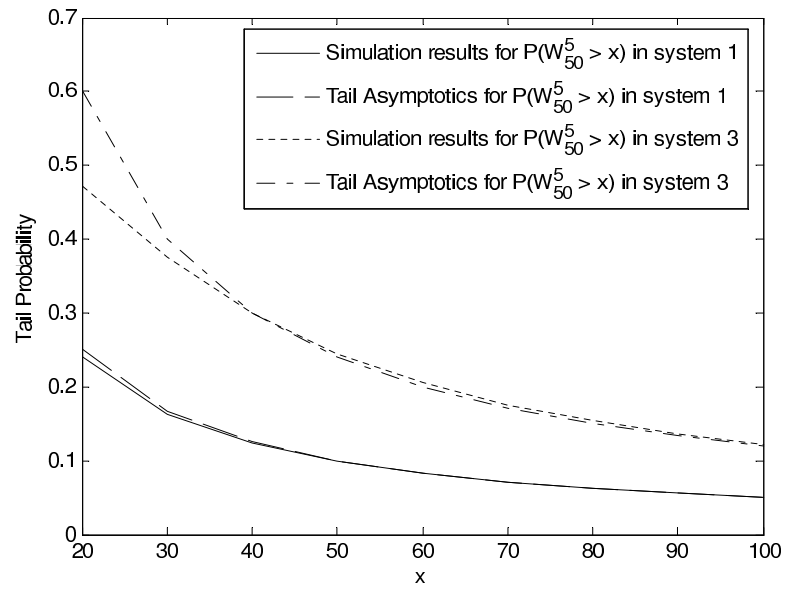
CHAPTER VII

CONTRIBUTION AND FUTURE RESEARCH DIRECTIONS

Motivated by data flows in telecommunication networks, there is an increasing interest in a variety of models with subexponential service times. In this thesis, we have investigated the tail asymptotics of transient and stationary cycle times and waiting times on a variety of queueing networks with FIFO service discipline and subexponential service time distributions. As Chapter 4 demonstrates, we were able to generalize the results of Ayhan, Palmowski, and Schlegel [9] to closed fork and join queues. In Chapter 5, we focused on closed K-stage tandem lines with finite buffers and subexponential service times under the manufacturing blocking and communication blocking schemes. We investigated the tail asymptotics of transient cycle times and waiting times. Also, we studied whether there exist conditions on service times such that tail asymptotics for transient characteristics also hold for their stationary counter parts. In Chapter 6, we considered open tandem queueing networks with subexponential service time distributions and finite buffers between stations. This system has K stations in tandem with infinite customers in front of first station and infinite room for finished customers after last station. We assume that this system is operating under the manufacturing blocking and communication blocking schemes. We analyzed the tail behavior of transient and stationary response times and waiting times. Moreover, we provided numerical experiments in order to study how fast the convergence of tail probabilities of key performance measures to their asymptotics counter parts in Chapters 4, 5, and 6. In the following, we suggest two future research topics related to our work.

Our first research direction is to generalize our results in Chapters 4 and 5 to a more general system which is the so-called closed (max,+) linear system. This system can be used to model various instances of queueing networks such as closed fork-join queues, tandem queueing networks with various kinds of blocking (manufacturing and communication), synchronized queueing networks etc. This system can be used to model window-based congestion control mechanism like TCP (Transmission Control Protocol), see Baccelli and Hong [14] for details. The dynamics of (max,+) linear systems can be captured by a simple stochastic difference equation from which one can obtain explicit expressions or bounds on system characteristics. In this system, we will focus on the tail characteristics of cycle times and waiting times. We have observed that the tail asymptotics of cycle times have the same structure on three different closed networks which are tandem queues with infinite buffers in Ayhan, Palmowski, and Schlegel [9], fork and joint networks with infinite buffers in Chapter 4, and tandem queues with finite buffers in Chapter 5, which all are the example of (max,+) linear systems. We have been motivated by this interesting results for the cycle times and thus, we will investigate whether the tail behavior of cycle times in general closed (max,+) linear systems could have the same structure. In addition, we will derive the general expression of the tail asymptotics of waiting times in general closed queueing networks. Finally, we will involve carrying out analysis to understand the convergence behavior of the tail asymptotics and consider testing the performance of the expressions for tail asymptotics in actual telecommunication systems.

The other future research direction is to extend the results in Chapter 6 to more general systems with finite buffers. For example, we can generalize our results to fork and join queues with an infinite number of customers in front of the first station and finite buffers between stations. Performance impact of transmission of multimedia has received a considerable amount of attention from academia and industry. It would therefore be interesting to study more complex systems which could be modeled as

fork and join queues. Our goal of this topic is to drive the expression of the tail asymptotics of response times and waiting times. Moreover, simulation studies could also be conducted to test the accuracy of the tail asymptotics of two key performance measures to their actual values.

REFERENCES

[1] ALTIOK, T. and STIDHAM, S., "A Note On Transfer Lines With Unreliable Machines, Random Processing Times, and Finite Buffers," *IIE Transactions*, vol. 14, no. 2, pp. 125–127, 1982.

[2] ARLITT, M. F. and WILLIAMSON, C. L., "Web server workload characterization: the search for invariants," *ACM SIGMETRICS Performance Evaluation Review*, vol. 24, no. 1, pp. 126–137, 1996.

[3] ASMUSSEN, S., "Subexponential Asymptotics for Stochastic Processes: Extremal Behavior, Stationary Distributions and First Passage Probabilities," *The Annals of Applied Probability*, vol. 8, no. 2, pp. 354–374, 1998.

[4] ASMUSSEN, S., HENRIKSEN, L. F., and KLÜPPELBERG, C., "Large claims approximations for risk processes in a Markovian environment," *Stochastic Processes and their Applications*, vol. 54, pp. 29–43, 1994.

[5] ASMUSSEN, S., KLÜPPELBERG, C., and SIGMAN, K., "Sampling at subexponential times, with queueing applications," *Stochastic Processes and their Applications*, vol. 79, no. 2, pp. 265–286, 1999.

[6] ASMUSSEN, S. and MØLLER, J. R., "Tail asymptotics for M/G/1 type queueing processes with subexponential increments," *Queueing Systems*, vol. 33, no. 1, pp. 153–176, 1999.

[7] ASMUSSEN, S., SCHMIDLI, H., and SCHMIDT, V., "Tail probabilities for non-standard risk and queueing processes with subexponential jumps," *Advances in Applied Probability*, vol. 31, pp. 422–447, 1999.

[8] AYHAN, H. and KIM, J. K., "A General Class of Closed Fork and Join Queues with Subexponential Service Times," *Stochastic Models*, vol. 23, no. 4, pp. 523–535, 2007.

[9] AYHAN, H., PALMOWSKI, Z., and SCHLEGEL, S., "Cyclic Queueing Networks with Subexponential Service Times," *Journal of Applied Probability*, vol. 41, no. 3, pp. 791–801, 2004.

[10] BACCELLI, F. and BRÉMAUD, P., *Elements of Queueing Theory: Palm Martingale Calculus and Stochastic Recurrences.* Springer, 2003.

[11] BACCELLI, F., COHEN, G. OLSDER, G. J., and QUADRAT, J.-P., *Synchronization and Linearity: An Algebra for Discrete Event Systems.* Wiley Series on Probability and Mathematical Statistics, 1992.

[12] BACCELLI, F. and FOSS, S., "Moments and tails in monotone-separable stochastic networks," *The Annals of Applied Probability*, vol. 14, no. 2, pp. 612–650, 2004.

[13] BACCELLI, F., FOSS, S., and LELARGE, M., "Tails in generalized Jackson networks with subexponential service-time distributions," *Journal of Applied Probability*, vol. 42, no. 2, pp. 513–530, 2005.

[14] BACCELLI, F. and HONG, D., "TCP is (max;+) Linear," in *Proceeding of ACM Sigcomm*, pp. 219–230, 2000.

[15] BACCELLI, F., LELARGE, M., and FOSS, S., "Asymptotics of Subexponential Max Plus Networks: the Stochastic Event Graph Case," *Queueing Systems*, vol. 46, no. 1, pp. 75–96, 2004.

[16] BACCELLI, F., SCHLEGEL, S., and SCHMIDT, V., "Asymptotics of stochastic networks with subexponential service times," *Queueing Systems*, vol. 33, no. 1, pp. 205–232, 1999.

[17] BALTRŪNAS, A., DALEY, D., and KLÜPPELBERG, C., "Tail behaviour of the busy period of a GI/GI/1 queue with subexponential service times," *Stochastic Processes and their Applications*, vol. 111, no. 2, pp. 237–258, 2004.

[18] BOROVKOV, A. A., "Stochastic processes in queueing theory," *Bulletin of the American Mathematical Society 83 (1977), 317-318.*, 1977.

[19] BORST, S., BOXMA, O., and JELENKOVIĆ, P., "Reduced-load equivalence and induced burstiness in GPS queues with long-tailed traffic flows," *Queueing Systems*, vol. 43, no. 4, pp. 273–306, 2003.

[20] BOXMA, O. and ZWART, B., "Tails in scheduling," *ACM SIGMETRICS Performance Evaluation Review*, vol. 34, no. 4, pp. 13–20, 2007.

[21] BRANDWAGN, A. and JOW, Y., "An approximation method for tandem queues with blocking," *Operations Research*, vol. 36, no. 1, pp. 73–83, 1988.

[22] COHEN, J. W., "Some Results on Regular Variation for Distributions in Queueing and Fluctuation Theory," *Journal of Applied Probability*, vol. 10, no. 2, pp. 343–353, 1973.

[23] CROVELLA, M. E. and BESTAVROS, A., "Self-similarity in World Wide Web traffic: evidence and possiblecauses," *Networking, IEEE/ACM Transactions on*, vol. 5, no. 6, pp. 835–846, 1997.

[24] CROVELLA, M. E. and LIPSKY, L., "Long-lasting Transient Conditions In Simulations With Heavy-tailed Workloads," *Simulation Conference Proceedings, 1997. Winter*, pp. 1005–1012, 1997.

[25] DENISOV, D., DIEKER, A., and SHNEER, V., "Large deviations for random walks under subexponentiality: the big-jump domain," *Ann. Probab*, vol. 36, no. 5, pp. 1946–1991, 2008.

[26] DENISOV, D., FOSS, S., and KORSHUNOV, D., "On lower limits and equivalences for distribution tails of randomly stopped sums," *Bernoulli*, vol. 14, no. 2, pp. 391–404, 2008.

[27] DIEKER, A. B. and LELARGE, M., "Tails for (max, plus) recursions under subexponentiality," *Queueing Systems*, vol. 53, no. 4, pp. 213–230, 2006.

[28] EMBRECHTS, P., KLÜPPELBERG, C., and MIKOSCH, T., *Modelling Extremal Events for Insurance and Finance*. Springer, 1997.

[29] EMBRECHTS, P. and VERAVERBEKE, N., "Estimates for the probability of ruin with special emphasis on the possibility of large claims," *Insurance: Mathematics and Economics*, vol. 1, no. 1, pp. 55–72, 1982.

[30] FELDMANN, A., GILBERT, A. C., WILLINGER, W., and KURTZ, T. G., "The changing nature of network traffic: scaling phenomena," *ACM SIGCOMM Computer Communication Review*, vol. 28, no. 2, pp. 5–29, 1998.

[31] FOSS, S., KONSTANTOPOULOS, T., and ZACHARY, S., "Discrete and Continuous Time Modulated Random Walks with Heavy-Tailed Increments," *Journal of Theoretical Probability*, vol. 20, no. 3, pp. 581–612, 2007.

[32] FOSS, S. and KORSHUNOV, D., "Heavy tails in multi-server queue," *Queueing Systems*, vol. 52, no. 1, pp. 31–48, 2006.

[33] FOSS, S., KORSHUNOV, D., and ZACHARY, S., "Convolutions of long-tailed and subexponential distributions," *Arxiv preprint arXiv:0810.1994*, 2008.

[34] Foss, S. and Richards, A., "On Sums of Conditionally Independent Subexponential Random Variables," *Arxiv preprint arXiv:0806.0490*, 2008.

[35] Fowler, T. B., "A Short Tutorial on Fractals and Internet Traffic," *The Telecommunications Review*, vol. 10, pp. 1–14, 1999.

[36] Geluk, J., "Some closure properties for subexponential distributions," *Statistics and Probability Letters*, 2009.

[37] Huang, T. and Sigman, K., "Steady-state asymptotics for tandem, split-match and other feedforward queues with heavy tailed service," *Queueing Systems*, vol. 33, no. 1, pp. 233–259, 1999.

[38] Jelenković, P. R., "Asymptotic analysis of queues with subexponential arrivals processes," *Self–Similar Network Traffic and Performance Evaluation, Eds. K. Park and W. Willinger, John Wiley & Sons, New York (NY)*, pp. 249–268, 2000.

[39] Jelenković, P. R. and Lazar, A. A., "Multiplexing on-off sources with subexponential on periods: Part I," *Proceedings IEEE INFOCOM*, vol. 1, pp. 187–195, 1997.

[40] Jelenković, P. R. and Lazar, A. A., "Subexponential Asymptotics of a Markov-Modulated Random Walk with Queueing Applications," *Journal of Applied Probability*, vol. 35, no. 2, pp. 325–347, 1998.

[41] Kim, J. K. and Ayhan, H., "Tandem Queues with Subexponential Service Times and Finite Buffers," *Queueing Systems submitted 2009.*

[42] Ko, B. and Tang, Q., "Sums of dependent nonnegative random variables with subexponential tails," *Journal of Applied Probability*, vol. 45, no. 1, pp. 85–94, 2008.

[43] Ko, S.-K. and Serfozo, R. F., "Response Times in M/M/s Fork-Join Networks," *Advances in Applied Probability*, vol. 36, no. 3, pp. 854–871, 2004.

[44] Leipus, R. and Šiaulys, J., "Asymptotic behaviour of the finite-time ruin probability under subexponential claim sizes," *Insurance Mathematics and Economics*, vol. 40, no. 3, pp. 498–508, 2007.

[45] Miyoshi, N., "On the Subexponential Properties in Stationary Single-server Queues: A Palm-martingale Approach," *Advances in Applied Probability*, vol. 36, no. 3, pp. 872–892, 2004.

[46] Pakes, A. G., "On the Tails of Waiting-Time Distributions," *Journal of Applied Probability*, vol. 12, no. 3, pp. 555–564, 1975.

[47] Paxson, V. and Floyd, S., "Wide area traffic: the failure of Poisson modeling," *Networking, IEEE/ACM Transactions on*, vol. 3, no. 3, pp. 226–244, 1995.

[48] Perros, H. and Altiok, T., "Approximate analysis of open networks of queues with blocking: Tandem configurations," *IEEE Transactions on Software Engineering*, vol. 12, no. 3, pp. 450–461, 1986.

[49] Resnick, S., *Heavy-tail phenomena: probabilistic and statistical modeling.* Springer, 2006.

[50] Scheller-Wolf, A. and Sigman, K., "Delay moments for FIFO GI/GI/s queues," *Queueing Systems*, vol. 25, no. 1, pp. 77–95, 1997.

[51] Shang, W., Liu, L., and Li, Q., "Tail asymptotics for the queue length in an M/G/1 retrial queue," *Queueing Systems*, vol. 52, no. 3, pp. 193–198, 2006.

[52] Takine, T., "Subexponential asymptotics of the waiting time distribution in a single-server queue with multiple Markovian arrival streams," *Stochastic Models*, vol. 17, no. 4, pp. 429–448, 2001.

[53] WHITT, W., "The impact of a heavy-tailed service-time distribution upon the M/GI/s waiting-time distribution," *Queueing Systems*, vol. 36, no. 1, pp. 71–87, 2000.

[54] WILLEKENS, E. and TEUGELS, J. L., "Asymptotic expansions for waiting time probabilities in an M/G/1 queue with long-tailed service time," *Queueing Systems*, vol. 10, no. 4, pp. 295–311, 1992.

[55] XIA, C. H., LIU, Z., SQUILLANTE, M. S., and ZHANG, L., "Lower Bounds for LRD/GI/1 Queues with Subexponential Service Times," *Probability in the Engineering and Informational Sciences*, vol. 18, no. 01, pp. 87–101, 2004.

[56] ZWART, A., "Tail asymptotics for the busy period in the GI/G/1 queue," *Mathematics of Operations Research*, pp. 485–493, 2001.