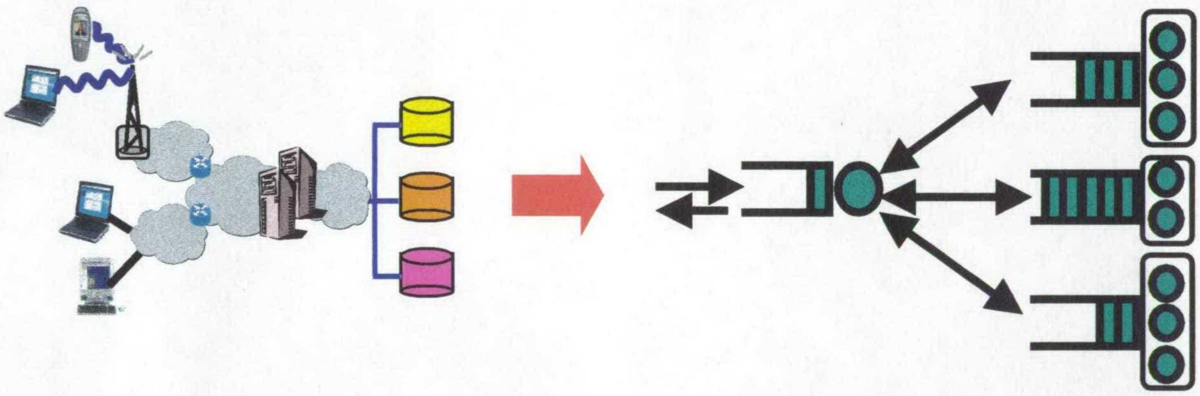
CWI

TILBURG UNIVERSITY

Approximations for Sojourn Times in Queueing Networks



Master's thesis
Karin van Wingerden
Amsterdam, February 2005

Approximations for Sojourn Times in Queueing Networks

Preface

This thesis is the result of a five-month internship at the *Centrum voor Wiskunde en Informatica* (CWI) in Amsterdam. It concludes my study Econometrics and Operations Research, in fact my Master Operations Research and Management Science, at *Tilburg University*. During my study and this internship, I have been supported by a number of people whom I would like to thank.

First of all, I would like to thank my supervisor at CWI, Rob van der Mei. He had always time to answer my questions. He assured me not to worry when I did not immediately understand everything, which was exactly what I needed! But there are more people who helped me with doing this research whom I would like to thank, like Bart Gijsen, Sindo Núñez Queija, Jacques Resing, and Wemke van der Weij. In particular, I would like to thank Onno Boxma, who, with his in-depth knowledge of queueing models, helped me on a regular basis. I would also like to thank my other colleagues at CWI for the nice atmosphere I experienced.

Furthermore, I would like to thank Hans Blanc, who supervised me during the internship on behalf of the university. I would also like to thank Dick den Hertog, who will be the chairman of my examination committee. Besides them, I would like to thank the rest of the staff of the Department *Econometrics and Operations Research*, together they created a pleasant study environment for me.

I want to thank Marcel for the time he spent on reading this thesis as an outsider.

Last but not least, I would like to thank my family and friends for the support they gave me during my study, but especially during these past five months. They stimulated me not to give up. Thanks for that!

Karin van Wingerden

Amsterdam, February 2005

Table of contents

Chapter 1

Introduction

1.1 Motivation

Information and communication technology (ICT) is becoming part of every day life. Nowadays, a lot of people do not even leave their house to get tickets for a concert or buy airline tickets, but simply order these on the Internet. Another well-known and frequently used application of ICT is electronic banking. Much of these kinds of Internet services are even accessible via the current generation mobile phones.

Example

A typical example of an Internet service accessible via mobile phones is the Local Weather Service (LWS). This service gives a mobile phone user the weather forecasts for the location where the user is at that moment. If the user sends a message to the LWS, he will receive a message with the weather forecasts after some amount of time. Between these two steps, sending a message (step 1 in figure 1.1) and receiving the information (step 5), some processes take place to make sure that the end user receives the correct information. First it has to be checked whether the user is authenticated to use this service (step 2); this can for example be checked in a database managed by the LWS service provider. Next, the location of the user has to be determined (step 3); therefore the X- and Y-coordinates of the user can be determined with some kind of a location service, managed by the end user's mobile service provider. Subsequently, the weather forecasts for the determined location and the right time are looked up in a database (step 4). The Dutch meteorological institute KNMI, for example, can offer such a weather service.

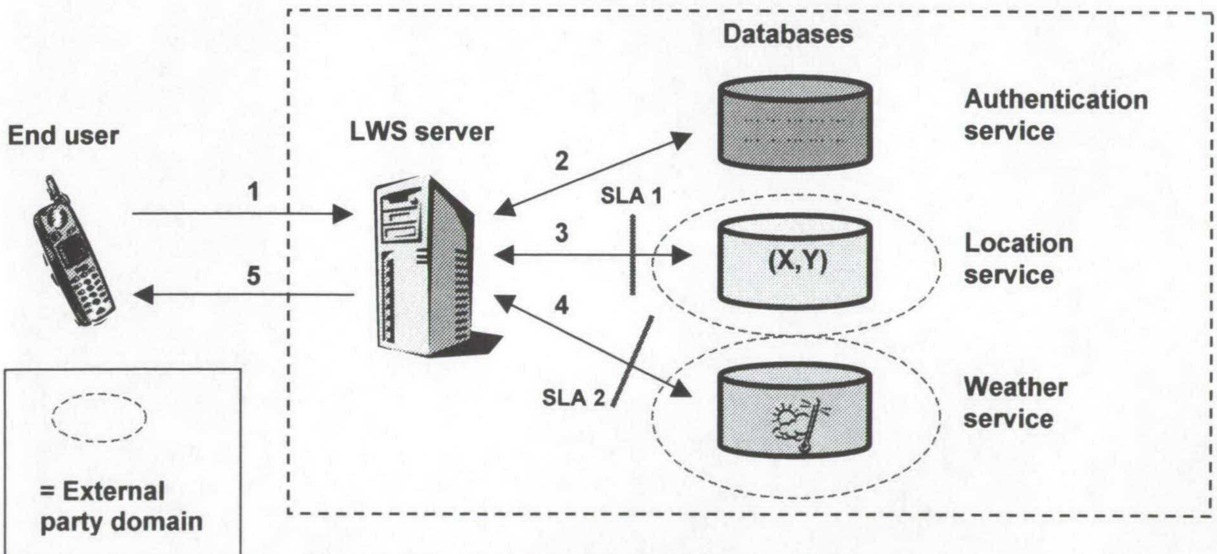


Figure 1.1 Access to Internet services by a mobile phone

For the commercial success of this kind of services, the ability to deliver an acceptable quality of service in terms of the end-to-end response time experienced by the end user is of key importance. The end-to-end response time depends on the performance over the domains managed by external parties, like the location service provider and the weather service provider in the example. To guarantee a given quality of service level the provider of the LWS-service needs to negotiate so-called Service Level Agreements (SLA's) with these external parties. An important question is which combination of SLA's leads to the preferred end-to-end quality of service.

Motivated by this, a key performance measure is the mean end-to-end response time experienced by the end user. This measure, however, does not give any information on what the probabilities are for a response time that is much higher than the average, or much lower. So, we are also interested in the variability of the response times.

Queueing model

We translate the problem described above into a performance model. Since delays play a dominant role, we will use the queueing network approach in this thesis. We assume that the reader has a background in queueing theory.

Our main interest is in the performance delivered by the LWS service provider (marked by the dotted area in figure 1.1). The additional delay in communication between the end user and the LWS service domain crossing the radio access network is beyond the scope of this study. The performance of the application server with significant server-side scripting is CPU-bound, i.e. its processing capacity is limited by the CPU-speed. This naturally leads to a processor sharing (PS) representation for the application server. In contrast, database access is typically I/O-bound. Database requests are handled one by one in order of arrival. This naturally leads to a model with multi-threaded first-come first-served (FCFS) queues, representing the databases.

These observations lead to the formulation of a queueing network as a performance model for the LWS-system. We analyse response times of transactions or database requests by modeling them as sojourn times of customers in an open queueing network. The sojourn time is the total time a customer spends in the network. In open networks, customers may arrive at any of the nodes, receive service at one or more nodes, and ultimately leave the network. The nodes represent servers.

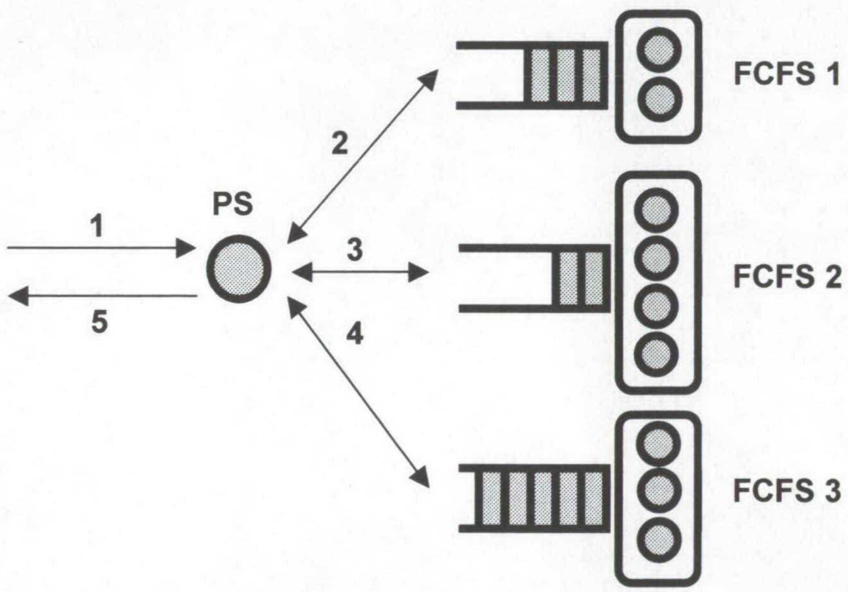


Figure 1.2 Queueing network under investigation

The queueing network corresponding to the LWS-system is given in figure 1.2. Note that the step numbers correspond to those in figure 1.1. The network consists of a single PS node, representing an application server, and multiple multi-server FCFS nodes, representing databases. In a PS node the server divides its capacity equally over all customers in the node and handles them simultaneously. A multi-server FCFS node consists of multiple parallel servers with the FCFS-discipline sharing a single queue of waiting customers.

1.2 Goal

The mean total sojourn time in the network is a measure for the average end-to-end response time experienced by the client. For some networks, an exact expression for the mean total sojourn time exists. We consider a number of variants of the queueing network in figure 1.2. The goal of this project is to derive exact expressions for the mean total sojourn time in the queueing networks considered if possible. If this is not possible for a network, then our goal is to develop explicit, fast-to-evaluate and accurate approximations for the mean total sojourn time in that network.

We are also interested in the variability of the end-to-end response times, which will be measured by the variance of the total sojourn time in the network. If an exact expression for the mean total sojourn time in a network exists, then our goal is to develop an explicit, fast-to-evaluate and accurate approximation for the variance of the total sojourn time in the network.

1.3 Overview

In chapter 2 we give an overview of literature about sojourn times in queueing networks, related to the models analysed in this thesis. We will use results from this literature in the following chapters.

In chapter 3 we consider a network with a single PS node and multiple multi-server FCFS nodes with Bernoulli feedback. The definition of Bernoulli feedback will be given in chapter 2. In this network the interarrival and all service times are exponentially distributed. We derive the closed-form expression for the mean total sojourn time in such a network and obtain an approximate expression for the variance of the total sojourn time in the network.

In chapter 4 we consider feedforward networks with a PS node and a single-server FCFS node with generally distributed service times at both nodes. The interarrival times are generally distributed as well. We approximate the mean sojourn times in this network by using the first and second moments of the service time distributions and interarrival time distributions.

In chapter 5 we consider a network with a PS node and a single-server FCFS node with Bernoulli feedback. The arrival process is a Poisson process, while the service times at both nodes are generally distributed. In this chapter the process of validating some of the results in Boxma et al. [8] is described and an improved approximation for the mean total sojourn time is given.

Finally, in chapter 6, we summarize the results of this investigation and we propose some possible extensions and topics for further research.

Chapter 2

Literature overview

2.1 Sojourn times in single nodes

2.1.1 PS node

Coffman, Muntz and Trotter [9] derived the transform of the sojourn time distribution for the M/M/1-PS node. Morrison [21] obtained an integral representation for this distribution. Ott [22] derived the sojourn time distribution for the M/G/1-PS case. Ramaswami [23] derived expressions for the first two moments of the sojourn time distribution for the GI/M/1-PS node; Jagerman and Sengupta [16] characterized the sojourn time distribution for the GI/M/1-PS node. Sengupta [24] obtained an approximation for the sojourn time distribution in the GI/G/1-PS node. We will use some of these results in the following chapters.

Van den Berg [4] derived the correlation between successive sojourn times at the M/G/1-PS node with feedback, i.e. after being served each customer either immediately arrives at the PS node again or departs permanently. We use this result in chapter 3 to approximate the variance of the total sojourn time in a feedback network with a single PS node and multiple FCFS nodes.

2.1.2 FCFS node

Many results are known for the sojourn time distribution in the FCFS queue. We concentrate on results for an FCFS queue with feedback. Takács [25] determined the Laplace-Stieltjes transform (LST) and the first two moments of a customer's total sojourn time in an M/G/1-FCFS queue with Bernoulli feedback, i.e. the probability of feedback does not depend on the number of completed services. Van den Berg [4] derived the joint distribution of the successive sojourn times of a customer in an M/M/1-FCFS queue with feedback. He gave expressions for the variance of the sojourn times in an FCFS node with Bernoulli feedback as well as for the variance of the sojourn times in an FCFS node with deterministic feedback.

2.2 Sojourn times in networks of queues

2.2.1 Product-form networks

Product-form networks are networks with a closed-form expression for the joint steady state distribution of the number of customers in each node. In 1957 J.R. Jackson [15] introduced the product-form distribution. His name is given to a class of product-form networks: Jackson networks. This class contains open networks, which fulfil the following conditions:

- The arrival process is a Poisson process, the arrival rate has to be independent of the number of customers at the nodes in the network;
- The service times are exponentially distributed, the service rate at a node may only depend on the number of customers in that particular node;
- The next node visited may depend on the present node, but has to be otherwise independent of the state of the system.

A more general class of product-form networks is the class of BCMP networks. Baskett, Chandy, Muntz, and Palacios [3] derived the joint distribution of queue sizes for a large class of networks. This class contains open, closed, and mixed networks of queues with different classes of customers. The queueing disciplines in the network may be FCFS and PS, but the network may also consist of lastcome first-served nodes (LCFS) or infinite server (IS) nodes. For the PS, LCFS, and IS service discipline the service times may be generally distributed, for the FCFS service discipline the service times must be exponentially distributed. Each customer belongs to a single class of customers while waiting or receiving service at a node, but may change classes and nodes according to fixed probabilities at the completion of a service request. External arrival processes must be Poisson processes.

For product-form networks, an exact expression for the mean sojourn time can be derived. Little's law, cf. [5], implies that the average time spent in a stable node by a customer is equal to the mean number of customers present in that node divided by the average number of customers that enter the node per time unit. Deriving higher order moments or tail probabilities for the sojourn time in a network is more difficult.

Boxma and Daduna [7] state that, in general, determining sojourn time distributions poses very complicated problems. An important exception is provided by the sojourn time distribution of a customer along a path in a product-form network, when this path satisfies certain overtake-free conditions. The possibility of customers, or their influences, overtaking each other, leads to dependencies that usually destroy any hope for an analytic solution for sojourn time distributions. Overtaking occurs when a customer physically overtakes another customer or when the influences generated by a customer overtake the customer. Overtaking can occur for two reasons. The first is overtaking due to the internal node structure, e.g. in a PS node it is possible that customers can leave before earlier arrived customers that have a larger service demand. The second reason is overtaking due to the structure of the network, e.g. in a non-acyclic network.

A network is called acyclic if a customer can never return to a station once visited. In a non-acyclic network, there can be two kinds of feedback: direct feedback, i.e. a customer immediately returns to a node after leaving it, and indirect feedback, i.e. a customer returns to a node after some time, e.g. after visiting another node. In the remainder of this thesis, we will also refer to direct feedback as short-circuiting. Indirect feedback will also be called delayed feedback. Feedback loops in a non-acyclic network imply dependent interarrival times at the nodes in the network. Hence, if the external arrival process at a certain node is a Poisson process and customers are fed back to that node, then the total arrival process in that node is not a Poisson process in general, not even if the service times are exponentially distributed.

Boxma and Daduna [7] also remark that the situation for non product-form networks is even worse; there are almost no explicit results. E.g., if service times are generally distributed, it is very difficult to find a closed-form expression for the mean sojourn time. In chapters 4 and 5 we derive approximations for the mean sojourn times in networks with generally distributed service times at the nodes.

2.2.2 Feedback queueing networks with a PS node and an FCFS node

Van der Mei et al. [20] study response times in a two-node queueing network with feedback. Their study is motivated by the performance analysis of response times in distributed information systems, where transactions are handled by iterative server and database actions. System response times are modelled as sojourn times in an open queueing network with a PS node and a single-server FCFS node. The network is shown in figure 2.1. External customers arrive at the PS node according to a Poisson process with rate λ. After receiving service at the PS node a customer proceeds to the FCFS node with probability p, and with probability $1-p$ the customer departs from the system. After a visit to the FCFS node, customers are fed back to the PS node. The service requirements at both nodes are exponentially distributed. The model is a product-form network, so closed-form expressions for the mean sojourn times in steady state exists. The variance of the sojourn times does not admit an exact expression; the complexity is caused by the possibility of overtaking, see section 2.2.1.

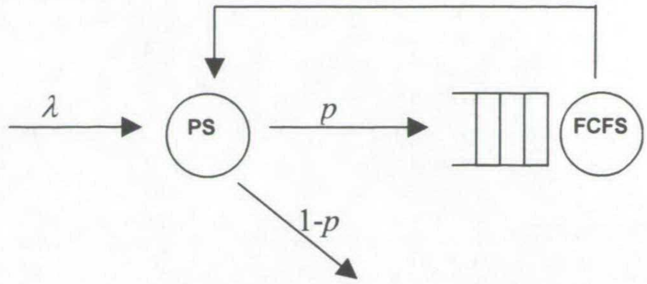


Figure 2.1 Feedback network with a PS node and a single-server FCFS node

The key assumptions for an approximation for the variance of the total sojourn time in this network are:

- The total arrival process at the PS node is a Poisson process with rate $\lambda/(1-p)$;
- The covariances between the successive sojourn times of a customer at the PS node in the network with delayed feedback are equal to those in a single M/M/1-PS node with direct feedback. Similarly, the covariances between the successive sojourn times of a customer at the FCFS node in the network with delayed feedback are equal to those in a single M/M/1-FCFS node with direct feedback;
- The total sojourn time at the first i visits to the PS node and the total sojourn time at the first j visits to the FCFS node are uncorrelated, for $i = 1,\ldots,N+1$ and $j = 1,\ldots,N$.

For non-acyclic queueing networks, the first assumption is known not to be true in general, see section 2.2.1. A result of the third assumption is that the covariance between the total sojourn time at the first i visits to the PS node and the total sojourn time at the first j visits to the FCFS node is assumed to be negligible. In general, also the third assumption is known to be not true, but observations suggest that the cross-correlation terms are rather small. The results of a variety of simulation experiments confirm the conjecture that the cross-correlation terms for the sojourn times of visits to different nodes are indeed negligible compared to correlation terms of successive visits to the same node.

With these assumptions a new methodology for deriving simple, explicit and fast-to-evaluate approximations for the variance of the sojourn times is proposed by Van der Mei et al. [20]. Numerical results demonstrate that the approximations are highly accurate in most model instances.

Boxma et al. [8] extend the model in [20]. The same model as described above is considered, but it is extended in two ways: general service times are allowed in both nodes and the authors present a more general approximation method that allows the approximation of sojourn time distributions while requiring somewhat less restrictive approximation assumptions. To determine approximations for the distribution of the total sojourn time the LST of the joint distribution of the total sojourn time in the PS node and the total sojourn time in the FCFS node is used.

The key approximation assumptions in [8] are:

- The total sojourn time at the first $j+1$ visits to the PS node is independent of the total sojourn time at the first j visits to the FCFS node, for $j = 1,2,\ldots$;
- The total sojourn time at the first j visits to the PS node has the same distribution as the sojourn time in the PS node short-circuited, i.e. with the FCFS node removed. Similarly, the total sojourn time at the first j visits to the FCFS node has the same distribution as the sojourn time in the FCFS node short-circuited, i.e. with the PS node removed.

The motivation for the first assumption is the same as for the third assumption in [20], as given above. Note that the third assumption in [20] is stronger: it suggests independence of all sojourn times at the various visits at different nodes of the network. Numerical experiments suggest that the second assumption works well when one is mainly interested in approximating the total sojourn time distribution, while one of the two queues has much larger mean sojourn times at each visit than the other one. In that case the second assumption should be quite accurate for that "bottleneck" queue. It is probably far from accurate for the other queue, but the contribution of the latter queue to the distribution of the total sojourn time is probably rather small. If the mean sojourn times at both queues are roughly equal, the authors suggest an improvement of the approximation. Instead of assuming that the total sojourn time to a node has the same distribution as the sojourn time in that node short-circuited, they propose to use an interpolation between the two extremes of short-circuiting and independence of successive sojourn times of a customer at the same queue.

The approximations in [8] are not yet validated. The process of validating and improving a part of the results from [8] is described in chapter 5 of this thesis.

2.2.3 Feedback queueing network with a PS node and multiple FCFS nodes

Gijsen et al. [14] study response times in a feedback queueing network with a single PS node and multiple single-server FCFS nodes. The network is another extension of the network in [20]. The motivation for this paper is the same as for [20], it can be found in section 2.2.2. Response times are modelled as sojourn times in an open queueing network with a PS node, representing an application server, and multiple FCFS nodes, all representing databases. The network is shown in figure 2.2. The service requirements at all nodes are exponentially distributed. External customers arrive at the PS node according to a Poisson process with rate λ. After receiving service at the PS node a customer proceeds either to FCFS-queue i with probability p_i, $0 \le p_i \le 1 \; \forall i = 1,...,M$, where M is the number of FCFS nodes in the network, or departs from the system with probability $1 - p$, $p = \sum_{i=1}^{M} p_i, 0 \le p \le 1$.

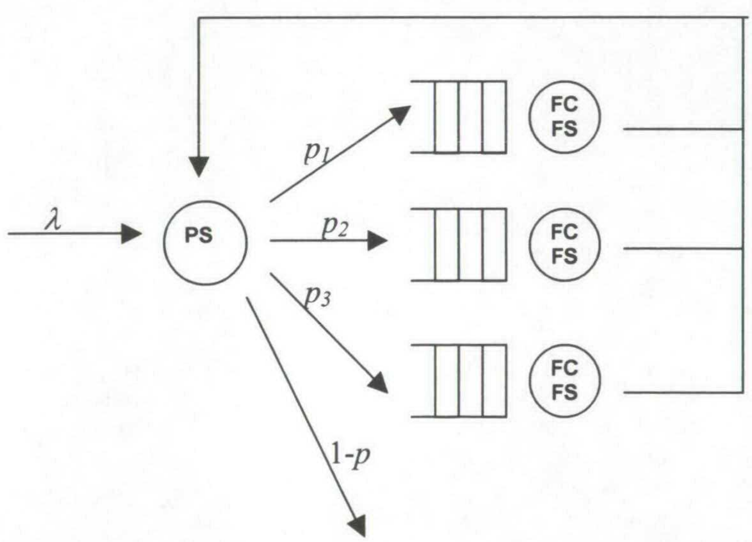


Figure 2.2 Feedback network with a single PS node and multiple multi-server FCFS nodes

The focus in [14] is on the mean and the variance of the sojourn time of an arbitrary customer in the system. The network is a product-form network, which immediately leads to a closed-form expression for the mean sojourn times. The variance of the sojourn times does not admit and exact expression, the complexity is caused by the possibility of overtaking, see section 2.2.1. Hence an approximation for the variance of the total sojourn time in the network is derived.

The key approximation assumptions in [14] are almost the same as in [20]:

- The total arrival process at the PS node is a Poisson process with rate $\lambda/(1-p)$;
- The covariances of the successive sojourn times of a customer at the PS node in the network with delayed feedback may be approximated by those in a single M/M/1-PS node with direct

feedback. Similarly, the covariances of the successive sojourn times of a customer at each FCFS node in the network with delayed feedback may be approximated by those in a single M/M/1-FCFS node with direct feedback;

- The total sojourn time at the first i visits to the PS node and the total sojourn time at the first j visits to the k^{th} FCFS node are uncorrelated, for $i = 1,..., N+1, j = 1,...,N$ and $k = 1,..., M$.

In general it is known that the first assumption is not true for non-acyclic networks, see section 2.2.1. The third assumption implies that the covariance between the total sojourn time at the first i visits to the PS node and the total sojourn time at the first j visits to the k^{th} FCFS node is assumed to be negligible. In general, the third assumption is known not to be true, see section 2.2.2. However, numerical results demonstrate that for queueing networks with multiple FCFS nodes the impact of ignoring cross-correlations on the approximation accuracy is even less than for queueing networks with a single FCFS node like in the previous section.

With these assumptions a new methodology for deriving explicit and fast-to-evaluate approximations for the variance of the sojourn times is proposed by Gijsen et al. [14]. Numerical results demonstrate that the approximations are highly accurate in most model instances. In chapter 3 we will extend these results for a network with multi-server instead of single-server FCFS nodes.

2.2.4 Queueing networks with multi-server FCFS nodes

Whitt [28] describes the software package Queueing Network Analyzer (QNA), which is developed to calculate approximate congestion measures for networks of queues. The first version of QNA analyses open networks of multi-server FCFS nodes and without capacity constraints. The QNA generates an approximation for the entire probability distribution for the steady state waiting time and approximates several other network performance measures. The external arrival processes need not to be Poisson and the service time distributions need not be exponential. This version of QNA uses two parameters to characterize the arrival processes and service times, one to describe the rate and the other to describe the variability. The nodes are then analysed as standard GI/G/m queues partially characterized by the first two moments of the interarrival time and service time distributions. Congestion measures for the network as a whole are obtained by assuming as an approximation that the nodes are stochastically independent given the approximate flow parameters.

The QNA is applicable for networks that satisfy the following requirements:

- The network is open;
- There are no capacity constraints, so there is no limit on the number of customers that can be in the entire network and each service facility has unlimited waiting space;
- There can be any number of servers at each node. These are identical independent servers, each serving one customer at a time;
- Customers are selected for service at each facility according to the FCFS discipline;
- There can be any number of customer classes, but customers cannot change classes;
- Customers can be created or combined at the nodes.

The general approach is to represent all the arrival processes and service time distributions by a few parameters. The congestion at each facility is then described by approximation formulas that depend only on these parameters. The parameters for internal flows are determined by applying elementary calculus that transforms the parameters for each of the three basic network operations: superposition or merging, thinning or splitting, and departure or flow through a queue. When the network is acyclic, the basic operations can be applied successively one at a time. In general, it is necessary to solve a system of equations or use an iterative method.

In [29] the performance of the QNA is described. QNA is compared with other approximations of several open networks of single-server queues and with simulations. The paper illustrates how to apply QNA and indicates the quality that can be expected from the approximations. Examples demonstrate the importance of the variability parameters used in QNA to describe non-Poisson arrival processes and non-exponential service time distributions.

In chapter 4 and 5 we will use Whitt's approximation for the mean waiting time in an FCFS queue with generally distributed interarrival and generally distributed service times. In these chapters, we will use his approximations for the squared coefficients of variation for the internal flows in case of superposition, splitting, and departure as well.

Chapter 3

A feedback queueing network with multi-server FCFS nodes

3.1 Introduction

In this chapter we extend the model considered by Gijsen et al. [14], see section 2.2.3, by using multi-server instead of single-server FCFS nodes. In the next section we will describe the model. In section 3.3 we will do the analysis, divided into results for the mean total sojourn time and results for the variance of the total sojourn time. In section 3.4, we discuss some numerical results.

3.2 Model description

First we introduce the notation used in this chapter.

Model input parameters

M	Number of FCFS nodes in the network;
c_k	Number of servers at the k^{th} FCFS node, $k = 1,\dots,M$;
p	Transition probability from the PS node to one of the FCFS nodes;
p_k	Transition probability from the PS node to the k^{th} FCFS node, $k = 1,\dots,M$;
q_k	Probability of feedback to the k^{th} FCFS node, $k = 1,\dots,M$;
p_{w_k}	Probability of waiting before receiving service at the k^{th} FCFS node, $k = 1,\dots,M$;
λ	External arrival rate;
λ_{PS}	Arrival rate at the PS node;
λ_{F_k}	Arrival rate at the k^{th} FCFS node, $k = 1,\dots,M$;
β_{PS}	Mean service time at the PS node;
β_{F_k}	Mean service time at the k^{th} FCFS node, $k = 1,\dots,M$;
ρ_{PS}	Load at the PS node;
ρ_{F_k}	Load at the k^{th} FCFS node, $k = 1,\dots,M$.

Random variables

N	Number of returns to the PS node;
N_k	Number of visits to the k^{th} FCFS node, $k = 1,\dots,M$;
L_{PS}	Stationary number of customers in the PS node;
L_{F_k}	Stationary number of customers in the k^{th} FCFS node, $k = 1,\dots,M$;
S_{PS}^i	Sojourn time at the i^{th} visit to the PS node, $i = 1,\dots,N$;
$S_{F_k}^i$	Sojourn time at the i^{th} visit to the k^{th} FCFS node, $i = 1,\dots,N_k$ and $k = 1,\dots,M$;
S_{PS}	Total sojourn time in the PS node;

S_{F_k} Total sojourn time in the k^{th} FCFS node, $k = 1,\ldots,M$;

S Total sojourn time in the network.

Here

$$p = \sum_{k=1}^{M} p_k \,, \tag{3.1}$$

$$q_k = \frac{p_k}{1 - p + p_k} \text{, cf. [6],} \tag{3.2}$$

$$p_{w_k} = \frac{{c_k}^{c_k} {\rho_{F_k}}^{c_k}}{c_k!} \left[1 + \sum_{n=1}^{c_k - 1} \left(1 - \frac{n}{c_k} \right) \frac{{c_k}^n {\rho_{F_k}}^n}{n!} \right]^{-1} \text{, cf. [5],} \tag{3.3}$$

and

$$N = \sum_{k=1}^{M} N_k \,. \tag{3.4}$$

Model

We consider a network consisting of a single PS node and M multi-server FCFS nodes. The service times at all nodes are exponentially distributed. External customers arrive at the PS node according to a Poisson process with rate λ. After departing from the PS node a customer proceeds to the k^{th} FCFS node with probability p_k, and with probability $1\text{-}p$ the customer departs from the system, with p as defined in (3.1). After each visit to any FCFS node customers are fed back to the PS node. An example of such a network is shown in figure 3.1.

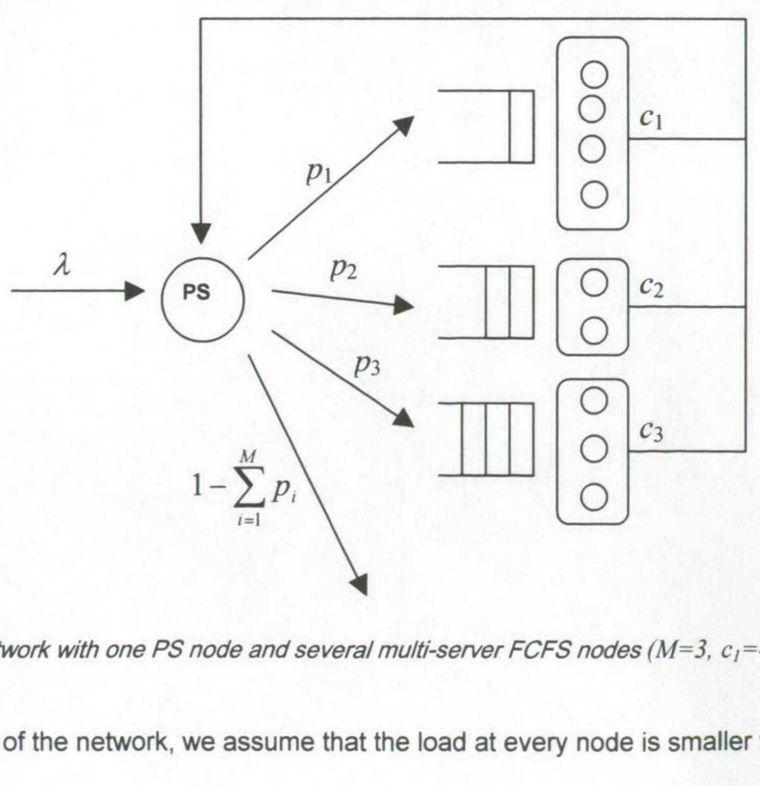


Figure 3.1 Network with one PS node and several multi-server FCFS nodes (M=3, c_1=4, c_2=2, c_3=3)

To ensure stability of the network, we assume that the load at every node is smaller than one.

Besides the external arrivals at the PS node with rate λ, there are also arrivals that are fed back with probability p. So the total arrival rate at the PS node is

$$\lambda_{PS} = \lambda + p\lambda + p^2\lambda + ... = \frac{\lambda}{1-p}. \tag{3.5}$$

The mean service time at the PS node is equal to β_{PS}. Hence it can be seen that the load at the PS node is given by

$$\rho_{PS} = \frac{\lambda\beta_{PS}}{1-p}. \tag{3.6}$$

The arrival rate at every FCFS node is equal to the probability that a customer goes to that FCFS node after receiving service at the PS node times the total arrival rate at the PS node

$$\lambda_{F_k} = p_k \frac{\lambda}{1-p}, k = 1,...,M. \tag{3.7}$$

Hence it can be seen that the load at the k^{th} FCFS node with mean service time β_{F_k} and c_k servers is given by

$$\rho_{F_k} = \frac{p_k}{1-p}\frac{\lambda\beta_{F_k}}{c_k}, k = 1,...,M. \tag{3.8}$$

The total sojourn time in the given network is defined as the sum of the sojourn times at all nodes in the network

$$S = S_{PS} + \sum_{k=1}^{M} S_{F_k} = \sum_{i=1}^{N+1} S_{PS}^i + \sum_{k=1}^{M}\sum_{i=1}^{N_k} S_{F_k}^i. \tag{3.9}$$

The sojourn time in a PS node depends on the number of customers that share the service capacity of that node. The sojourn time in an FCFS node consists of the waiting time in the queue plus the service time. The number of visits to the PS node is always equal to the number of visits to the FCFS nodes plus one, because after service completion at one of the FCFS nodes a customer always returns to the PS node before eventually leaving the system.

3.3 Analysis

3.3.1 Exact result for mean total sojourn time

Similar to the networks in [14] and [20], the network in this chapter is a product-form network. Therefore we can derive an exact result for the mean total sojourn time in the network described in section 3.2. Below we apply Little's law to every node in the network, so that we can calculate a customer's mean sojourn time at each node separately.

The stationary number of customers in the PS node, L_{PS}, is geometrically distributed, i.e. $P\{L_{PS} = l\} = (1-\rho_{PS})\rho_{PS}{}^l, l = 0,1,2,...$. Hence the mean number of customers in the PS node is

$$E\{L_{PS}\} = \frac{\rho_{PS}}{1-\rho_{PS}}. \tag{3.10}$$

Now we can derive the mean sojourn time for the i^{th} visit to the PS node by Little's law

$$E\{S_{PS}{}^{i}\} = \frac{E\{L_{PS}\}}{\lambda_{PS}} = \frac{\beta_{PS}}{1-\rho_{PS}},\ i = 1,\dots,N. \tag{3.11}$$

The mean number of customers in each multi-server FCFS node is, cf. [5],

$$E\{L_{F_k}\} = c_k \rho_{F_k} + \frac{\rho_{F_k} p_{W_k}}{1-\rho_{F_k}},\ k = 1,\dots,M, \tag{3.12}$$

where p_{w_k} as given in (3.3).

Now we can derive the mean sojourn time for the i^{th} visit to the k^{th} FCFS node by Little's law

$$E\{S_{F_k}{}^{i}\} = \frac{E\{L_{Fk}\}}{\lambda_{F_k}} = \beta_{F_k} + \frac{p_{w_k}}{c_k}\frac{\beta_{F_k}}{1-\rho_{F_k}},\ i = 1,\dots,N_k,\ k = 1,\dots,M. \tag{3.13}$$

The number of times a customer is fed back to the PS node, N, is easily seen to be geometrically distributed, i.e. $P\{N = n\} = (1-p)p^n$, $n = 0,1,2,\dots$. Hence

$$E\{N\} = \frac{p}{1-p}. \tag{3.14}$$

Similarly, the number of times a customer is fed back to the k^{th} FCFS node, N_k, is easily seen to be geometrically distributed, i.e. $P\{N_k = n_k\} = (1-q_k)q_k{}^{n_k}$, $n_k = 0,1,\dots,$ for $k = 1,\dots M$. Hence

$$E\{N_k\} = \frac{q_k}{1-q_k} = \frac{p_k}{1-p},\ k = 1,\dots,M. \tag{3.15}$$

We obtain the following closed-form expression for the mean total sojourn time of an arbitrary customer

$$\begin{aligned} E\{S\} &= E\left\{\sum_{i=1}^{N+1} S_{PS}^{i} + \sum_{k=1}^{M}\sum_{i=1}^{N_k} S_{F_k}^{i}\right\} = \left(E\{N\}+1\right)E\{S_{PS}^{1}\} + \sum_{k=1}^{M} E\{N_k\}E\{S_{F_k}^{1}\} \\ &= \frac{1}{1-p}\frac{\beta_{PS}}{1-\rho_{PS}} + \sum_{k=1}^{M}\frac{p_k}{1-p}\left(\beta_{F_k} + \frac{p_{w_k}}{c_k}\frac{\beta_{F_k}}{1-\rho_{F_k}}\right). \end{aligned} \tag{3.16}$$

The first equation follows by the definition of the total sojourn time (3.9). The second equation follows by applying Wald's equation, see [26]. The last equation is obtained by substitution of (3.11), (3.13), (3.14), and (3.15).

3.3.2 Approximation for variance of total sojourn time

Boxma and Daduna [7] obtain an expression for the joint distribution of the sojourn time in a product-form network under certain conditions. One of the conditions is that the paths the customers have to traverse have the overtake-free property. However, in the network in this chapter overtaking can occur. Overtaking introduces correlation between the sojourn times of customer visits at the nodes in the queueing network. Instead of determining an exact result for the variance of the total sojourn time in our network, we will develop an approximation for it.

Note that the successive sojourn times of a customer at the same node are identically distributed. This can be seen as follows. The model under consideration is a multi-class product-form network, where the customer classes are defined as follows. Each customer enters the network at the PS node as a class-0 customer, and its class number is incremented from i to i +1 any time the customers jumps from one node to the next. Then for each customer its class number indicates the number of node visits since its arrival in the system. According to the Arrival Theorem (cf., e.g., theorem 4.1 in [5]) at the instant when a customer jumps from one node to another the joint queue-length distribution of the network is equal to the stationary joint queue-length distribution of the same network but without the jumping customer. Note that this is regardless of its class number. This implies that the successive sojourn times of a customer at the same node are identically distributed.

In appendix B the complete derivation of the approximation for the variance of the total sojourn time in the network described in section 3.2 is given. In this section we will just write down the main steps and results. We start with a general formula for the variance of the total sojourn time by substituting the definition of the total sojourn time in the network

$$Var\{S\} = Var\left\{\sum_{i=1}^{N+1} S_{PS}^i + \sum_{k=1}^{M}\sum_{i=1}^{N_k} S_{F_k}^i\right\}. \tag{3.17}$$

In appendix B we obtain the following general expression for the variance of the total sojourn time in the network described in section 3.2:

$$\begin{aligned}
Var\{S\} = &\sum_{n=0}^{\infty}(n+1)Var\{S_{PS}^1\}(1-p)p^n + \sum_{n=0}^{\infty}\sum_{i\neq j} Cov\{S_{PS}^i, S_{PS}^j\}(1-p)p^n + \\
&\sum_{k=1}^{M}\sum_{n_k=0}^{\infty} n_k Var\{S_{F_k}{}^1\}(1-q_k){q_k}^{n_k} + \sum_{k=1}^{M}\sum_{n_k=0}^{\infty}\sum_{i\neq j} Cov\{S_{F_k}^i, S_{F_k}^j\}(1-q_k){q_k}^{n_k} + \\
&2\sum_{k=1}^{M}\sum_{n_1=0}^{\infty}\cdots\sum_{n_M=0}^{\infty} Cov\left\{\sum_{i=1}^{n+1} S_{PS}^i, \sum_{j=1}^{n_k} S_{F_k}^j\right\} f(n_1,\ldots,n_M) + \\
&\sum_{k\neq m}\sum_{n_1=0}^{\infty}\cdots\sum_{n_M=0}^{\infty} Cov\left\{\sum_{j=1}^{n_k} S_{F_k}^i, \sum_{j=1}^{n_m} S_{F_m}^i\right\} f(n_1,\ldots,n_M) + \\
&Var\left\{\sum_{k=1}^{M} N_k\left(\{S_{PS}^1\} + E\{S_{F_k}^1\}\right)\right\},
\end{aligned} \tag{3.18}$$

where $f(n_1,\ldots,n_M)$ denotes the joint probability distribution of the number of visits to the FCFS nodes.

By ignoring the dependencies overtaking can create in the network, we ignore the correlations between sojourn times. The result is a simple approximation for the variance of the total sojourn time in the network considered in this chapter. Because we do not take into account the correlations between sojourn times, we expect this simple approximation to underestimate the variance of the total sojourn time. This approximation will be used to compare with the final approximation, which follows under less restrictive assumptions.

$$\begin{aligned}
Var_{simple}\{S\} \approx &\frac{1}{1-p}\frac{2+\rho_{PS}}{2-\rho_{PS}}\left(\frac{\beta_{PS}}{1-\rho_{PS}}\right)^2 + \sum_{k=1}^{M}\frac{q_k}{1-q_k}\left({\beta_{Fk}}^2 + \frac{p_w(2-p_w){\beta_{Fk}}^2}{{c_k}^2(1-\rho_{Fk})^2}\right) \\
&+ \sum_{k=1}^{M}\frac{q_k}{(1-q_k)^2}\left(\frac{\beta_{PS}}{1-\rho_{PS}} + \beta_{Fk} + \frac{p_{w_k}\beta_{Fk}}{c_k(1-\rho_{Fk})}\right)^2.
\end{aligned} \tag{3.19}$$

For a refined approximation for the variance of the total sojourn time in the network, we start with the last term in (3.18). Since the number of visits to the k^{th} FCFS node, N_k, is geometrically distributed with parameter q_k, we know that

$$\sum_{k=1}^{M} Var\{N_k\} = \sum_{k=1}^{M} \frac{q_k}{(1-q_k)^2}. \tag{3.20}$$

With some calculus, cf. [13], we find that

$$\sum_{k \neq m} Cov\{N_k, N_m\} = Var\{N\} - Var\left\{\sum_{k=1}^{M} N_k\right\} = \sum_{k \neq m} \frac{p_k p_m}{(1-p)^2}. \tag{3.21}$$

By substitution, we obtain

$$\begin{aligned} Var\left\{\sum_{k=1}^{M} N_k \left(E\{S_{PS}^1\} + E\{S_{F_k}^1\}\right)\right\} = & \sum_{k=1}^{M} \frac{q_k}{(1-q_k)^2}\left(E\{S_{PS}^1\} + E\{S_{F_k}^1\}\right)^2 + \\ & \sum_{k \neq m} \frac{p_k p_m}{(1-p)^2}\left(E\{S_{PS}^1\} + E\{S_{F_k}^1\}\right)\left(E\{S_{PS}^1\} + E\{S_{F_m}^1\}\right). \end{aligned} \tag{3.22}$$

The expressions for $E\{S_{PS}{}^1\}$ and $E\{S_{F_k}{}^1\}$ are given in (3.11) and (3.13) respectively.

Now we still need formulas for $Var\{S_{PS}^1\}, Var\{S_{F_k}^1\}, Cov\{S_{PS}^i, S_{PS}^j\}, Cov\{S_{F_k}^i, S_{F_k}^j\}, Cov\{S_{PS}^i, S_{F_k}^j\}$ and $Cov\{S_{F_k}^i, S_{Fm}^i\}$, for any $i, j = 1,2,\ldots$, $k = 1,\ldots, M$, and $k \neq m$.

Therefore we use the following approximation assumptions:

1 The arrival process at the PS node is a Poisson process with parameter $\lambda/(1-p)$.
2 The covariances of the successive sojourn times of a customer at the PS node in the network with delayed feedback may be approximated by those in a single M/M/1-PS node with direct feedback. Similarly, the covariances of the successive sojourn times of a customer at the FCFS node in the network with delayed feedback may be approximated by those in a single M/M/1-FCFS node with direct feedback.
3 The total sojourn time at the first i visits to the PS node and the total sojourn time at the first j visits to the k^{th} FCFS node are uncorrelated, for $i = 1,\ldots, N+1$, $j = 1,\ldots,N$ and $k = 1,\ldots, M$.
4 Sojourn times of a customer at different FCFS nodes are uncorrelated.

Approximation assumption 1

In general, it is known that the first approximation assumption is not true for the network under consideration, due to the possibility of feedback. The feedback loop implies dependent interarrival times at the nodes, which results in violation of the Poisson assumption. However, we need the first assumption to approximate the variance of the sojourn time in the PS node. Under this approximation assumption we can use Ott's result, cf. [22], for the variance of the sojourn time in an M/M/1-PS node:

$$Var\{S_{PS}^1\} \approx \frac{2+\rho_{PS}}{2-\rho_{PS}}\left(\frac{\beta_{PS}}{1-\rho_{PS}}\right)^2. \tag{3.23}$$

If we assume that the arrival process at the PS node is a Poison process and all service times are exponentially distributed, then the arrival process at the FCFS node is a Poisson process as well. We derive the variance of the sojourn time in the k^{th} FCFS node by using the theory about M/M/c queues in [5]

$$Var\{S_{F_k}^1\} \approx \beta_{Fk}^{\ 2} + \frac{p_w(2-p_w)\beta_{Fk}^{\ 2}}{c_k^{\ 2}(1-\rho_{Fk})^2}, k = 1,...,M. \tag{3.24}$$

Approximation assumption 2

Van den Berg [4] derives exact expressions for the covariances of the successive sojourn times for single-server FCFS and PS queues with direct feedback, where customers upon service completion are immediately fed back into the system with a given probability. Our model, instead, has a delayed feedback mechanism: upon departing from the PS node, a customer is first processed by the FCFS node (if not leaving the system immediately) before returning to the PS node. Similarly, after leaving any FCFS node, a customer is first processed at least once by the PS node before returning to the same FCFS node. Based on the second approximation assumption, we approximate the covariances between the successive sojourn times at the same node by the exact results in [4] for systems with direct feedback. In [27] the covariances of the successive sojourn times for single-server FCFS and PS queues are derived from results in [4], we apply the results from [27].

The covariance between the successive sojourn times at the PS node is given by (4.25) in [27]

$$Cov\{S_{PS}^i, S_{PS}^{i+l}\} \approx \frac{\rho_{PS}\beta_{PS}^{\ 2}}{(1-\rho_{PS})^2(2-\rho_{PS}-p+\rho_{PS}p)^{l+1}}, i = 1,2,..., l = 1,2,.... \tag{3.25}$$

For the covariance between the successive sojourn times at the multi-server FCFS nodes we extend (4.26) in [27] in a way that is given in appendix B. We derive the following approximate expression

$$Cov\{S_{F_k}^i, S_{F_k}^{i+l}\} \approx p_{w_k}\left(\rho_{F_k}(1-q_k)+q_k\right)^{l-1}\left(\beta_{Fk}^{\ 2} + p_{wk}\left(2-p_{wk}\right)\frac{\beta_{Fk}^{\ 2}}{c_k^{\ 2}(1-\rho_{Fk})^2}\right), \tag{3.26}$$

for $i = 1,2,...,\ k = 1,2,...,M$, and $l = 1,2,....$

Approximation assumption 3

In general, the third approximation assumption is known not to be true. However, the product-form solution for the present model, see Disney and Koenig [11], implies that the numbers of customers at the two nodes are independent in equilibrium. Gijsen et al. [14] remark that the sojourn time in the FCFS nodes is closely related to the number of customers at that node in the following way. If a customer finds n_{Fk} customers at the k^{th} FCFS node upon arrival, then the sojourn time simply consists of $n_{Fk}+1$ independent successive exponential phases each with rate β_{F_k}. This results in an Erlang distribution with shape parameter $n_{Fk}+1$ and rate parameter β_{F_k}. For the PS node, the correlation between the sojourn times and number of customers present upon arrival is less clear, and intuitively seems to be weaker than for FCFS nodes. These observations suggest that the cross-correlation terms are rather small. Van der Mei et al. [20] found that the cross-correlation coefficients (between PS and FCFS nodes) were about a factor two smaller than the correlation coefficient for successive sojourn times at the PS node. They also found that the correlation coefficient for successive sojourn times at the FCFS node were about three times larger than the PS node correlation coefficient. These results confirm the conjecture that the cross-correlation terms for the sojourn times of visits to different nodes are indeed negligible compared to the correlation terms of successive visits to the same node. Gijnen et al. [14] denote that for queueing networks with several FCFS nodes the impact of ignoring cross-correlations on the approximation accuracy is even less.

The third approximation assumption implies that

$$Cov\{S_{PS}^{i}, S_{F_k}^{j}\} \approx 0, i \neq j. \tag{3.27}$$

Approximation assumption 4

As for the third approximation assumption, we know that the fourth approximation assumption is not true in general. However, for the same reasons as denoted for ignoring the cross-correlations between the PS node and FCFS nodes, we will also ignore the cross-correlations between two different FCFS nodes. Finally, the fourth approximation assumption implies that

$$Cov\{S_{F_k}^{i}, S_{F_m}^{i}\} \approx 0, i = 1,2,\ldots, k \neq m. \tag{3.28}$$

Final approximation

By substitution of equations (3.22) to (3.28), we finally obtain the following approximation for the variance of the total sojourn time in a feedback network with a single PS node and several multi-server FCFS nodes

$$\begin{aligned} Var\{S\} \approx {} & \frac{1}{1-p}\frac{2+\rho_{PS}}{2-\rho_{PS}}\left(\frac{\beta_{PS}}{1-\rho_{PS}}\right)^2 + \frac{2p\rho_{PS}\beta_{PS}^{\;2}}{(1-\rho_{PS})^2(1-p)^2(2-\rho_{PS}-p+p\rho_{PS})(2-\rho_{PS})} \\ & + \sum_{k=1}^{M}\frac{q_k}{1-q_k}\left(\beta_{Fk}^{\;2} + \frac{p_w(2-p_w)\beta_{Fk}^{\;2}}{c_k^{\;2}(1-\rho_{Fk})^2}\right) + \sum_{k=1}^{M}\frac{2q_k^{\;2}p_{w_k}\beta_{F_k}^{\;2}((1-\rho_{F_k})^2c_k^{\;2}+p_{w_k}(2-p_{w_k}))}{(1-q_k)^2(1-\rho_{F_k})^2(1-q_k\rho_{F_k}+q_k)c_k^{\;2}} \\ & + \sum_{k=1}^{M}\frac{q_k}{(1-q_k)^2}\left(\frac{\beta_{PS}}{1-\rho_{PS}} + \beta_{Fk} + \frac{p_{w_k}\beta_{Fk}}{c_k(1-\rho_{Fk})}\right)^2 \\ & + \sum_{k \neq m}\frac{p_k p_m}{(1-p)^2}\left(\frac{\beta_{PS}}{1-\rho_{PS}} + \beta_{Fk} + \frac{p_{w_k}\beta_{Fk}}{c_k(1-\rho_{Fk})}\right)\left(\frac{\beta_{PS}}{1-\rho_{PS}} + \beta_{Fm} + \frac{p_{w_m}\beta_{Fm}}{c_m(1-\rho_{Fm})}\right). \end{aligned} \tag{3.29}$$

As mentioned before, the details of this derivation can be found in appendix B.

3.4 Numerical results

To assess the accuracy of the approximations for the variance of the sojourn times proposed in section 3.3.2, we have performed numerous numerical experiments, comparing the approximations with simulations. For every simulation the number of multi-server FCFS nodes is taken equal to three. We have validated the accuracy of the approximations for many parameter combinations, by varying the mean service times at all nodes, the number of servers in the FCFS nodes and the values of p_1, p_2 and p_3. More details about the simulations can be found in appendix A.

We calculated the point estimates for the variance of the sojourn times, and its 95% confidence intervals. By comparing the point estimates based on the simulations with the approximations, we calculated the relative error of the approximations in the following way

$$\Delta\% = \frac{approximation - simulation}{simulation} \times 100, \tag{3.30}$$

where $\Delta\%$ is the relative error, *approximation* is the approximated value of the variance of the total sojourn time in the network, and *simulation* is the point estimation of the variance of the total sojourn time in the network based on the simulations.

To show that the quite complex, but closed-form approximation given in section 3.3.2 indeed leads to a higher level of accuracy than simpler variants of approximations for the variance of the total sojourn time in the network, we compare our approximation (3.29) with the simple variant given in (3.19).

In the tables below we denote the approximated value of the variance of the total sojourn time (3.29) in the network by "appr", the results of the simple approximation (3.19) by "simple", the point estimation of the variance of the total sojourn time in the network based on the simulations by "sim", and the 95% confidence interval by "95% c.i.".

We considered the feedback probabilities, the loads at the nodes, and the number of servers at the FCFS nodes as given in the tables. For those parameter values we calculated the corresponding mean service times $\beta_{F_1}, \beta_{F_2}, \beta_{F_3}$ and β_{PS}. In tables 3.1 to 3.5 the first three digits of the values for the mean service times are shown.

Table 3.1 shows the results for the first case, networks with three identical single-server FCFS nodes and equal probabilities for visiting each of the FCFS nodes. Furthermore we fixed the external arrival rate. We varied the feedback probability p and the loads at the nodes ρ_{PS} and ρ_{F_i}. For every run in table 3.1 it holds that $\lambda = 1, p_1 = p_2 = p_3 = p/3, \beta_{F_1} = \beta_{F_2} = \beta_{F_3}, c_1 = c_2 = c_3 = 1$ and $\rho_{PS} = \rho_{F_1} = \rho_{F_2} = \rho_{F_3}$.

p	β_{PS}	β_{F_i}	ρ_{PS}	ρ_{F_i}	sim	95% c.i.	appr	$\Delta\%$	simple	$\Delta\%$
0.3	0.14	1.40	0.20	0.20	3.72	(3.69, 3.74)	3.71	-0,29	3.18	-14.40
0.3	0.35	3.50	0.50	0.50	60.80	(59.43, 62.16)	61.44	1,05	51.21	-15.78
0.3	0.56	5.60	0.80	0.80	997.59	(926.70, 1068.48)	1023.47	2,59	826.77	-17.12
0.6	0.08	0.40	0.20	0.20	1.81	(1.80, 1.83)	1.82	0,38	1.22	-33.00
0.6	0.20	1.00	0.50	0.50	31.18	(30.68, 31.68)	31.17	-0,02	19.63	-37.05
0.6	0.32	1.60	0.80	0.80	541.74	(519.20, 564.29	540.39	-0,25	318.29	-41.25
0.9	0.02	0.07	0.20	0.20	1.18	(1.18,1.19)	1.18	-0.01	0.49	-58.43
0.9	0.05	0.17	0.50	0.50	20.88	(20.60,21.16)	20.82	-0.29	7.93	-62.04
0.9	0.08	0.27	0.50	0.50	273.42	(262.59,284.26)	268.42	-1.83	90.13	-67.04
0.9	0.08	0.27	0.80	0.80	369.17	(348.97,389.36)	375.74	1.78	127.89	-65.36

Table 3.1 Variance of the total sojourn time for a network with three symmetric single-server FCFS nodes: approximations versus simulations

The results presented in table 3.1 show that the approximations are highly accurate for all parameter combinations considered, with a worst-case scenario of only 2.59%. Further, the results show that the approximation is much more accurate than the simple approximation, which shows errors up to 67%! Our approximations are within the 95%-confidence interval in all cases. Note that the approximation does not always underestimate or overestimate the simulated variance. In fact, table 3.1 just demonstrates that our approximation is accurate for a network with a single PS node and multiple single-server FCFS nodes.

In table 3.2 we show results for a network with multi-server FCFS nodes. This second case is a symmetrical network again: three identical FCFS servers, equal probabilities for visiting each FCFS node, and fixed external arrival rate, $\lambda = 1$. For every run in table 3.2 it holds that $p_1 = p_2 = p_3 = 0.3, \beta_{F_1} = \beta_{F_2} = \beta_{F_3}, c_1 = c_2 = c_3 > 1$ and $\rho_{F_1} = \rho_{F_2} = \rho_{F_3}$.

Unfortunately there is no space for confidence intervals in table 3.2 and the following tables. For every run we checked if our approximation is in the 95% confidence interval, and it is in most cases. It is remarkable that for the cases where the approximation is not in the 95% confidence interval, it is always an overestimation of the simulation value. Apparently, the approximation always overestimates the variance of the sojourn time for the cases we considered. This is probably due to the interpolation used to approximate the covariance between the sojourn times for successive visits to the same multi-server FCFS node. Nevertheless the approximation is very close to the simulation value in all cases considered.

The results presented in table 3.2 show that our approximation works very well for networks with multi-server FCFS nodes as well. Again, it can be seen that our approximation is much more accurate than the simple one.

β_{PS}	β_{F_i}	c_i	ρ_{PS}	ρ_{F_i}	sim	appr	$\Delta\%$	simple	$\Delta\%$
0.02	0.33	2	0.20	0.50	22.90	23.67	3.35	9.50	-58.54
0.02	0.53	2	0.20	0.80	271.43	269.03	-0.88	93.56	-65.53
0.05	0.13	2	0.50	0.20	5.88	6.08	3.53	2.39	-59.34
0.05	0.53	2	0.50	0.80	306.54	306.97	0.14	107.43	-64.95
0.08	0.13	2	0.80	0.20	47.85	48.90	2.20	14.36	-70.00
0.08	0.33	2	0.80	0.50	87.11	90.55	3.95	30.96	-64.46
0.01	0.33	4	0.10	0.25	11.85	11.90	0.46	5.33	-54.99
0.01	1.20	4	0.10	0.90	1680.76	1637.24	-2.59	550.25	-67.26
0.03	0.13	4	0.25	0.10	2.67	2.70	1.18	1.17	-56.02
0.08	1.20	4	0.75	0.90	1829.87	1842.08	0.67	629.58	-65.59
0.09	0.13	4	0.90	0.10	230.90	235.73	2.09	59.96	-74.03
0.09	1.00	4	0.90	0.75	692.27	718.54	3.79	246.18	-64.44
0.02	1.67	10	0.20	0.50	286.38	290.27	1.36	129.57	-54.75
0.05	0.67	10	0.50	0.20	56.97	57.64	1.17	25.33	-55.55
0.08	0.67	10	0.80	0.20	123.59	128.96	4.34	48.69	-60.60
0.08	1.67	10	0.80	0.50	430.20	440.47	2.39	184.36	-57.15

Table 3.2 Variance of the total sojourn time for a network with three symmetric multi-server FCFS nodes: approximations versus simulations

For the results in tables 3.1 and 3.2 it is assumed that the three FCFS servers are identical. The tables show that our approximation works very well in the given symmetric cases. To investigate the impact of asymmetry in the nodes on the accuracy of the approximations, we have also considered a variety of parameter combinations with more asymmetric characteristics.

In table 3.3 we show the results for three FCFS nodes, with the same number of servers at each of the nodes. In this case, the probabilities of visiting each of the nodes as well as the service times at the FCFS nodes are not equal for all FCFS nodes; in such a way that the loads at the different FCFS nodes are equal. For every case in table 3.3 it holds that $\lambda = 1$, $p_1 \neq p_2 \neq p_3$, $\beta_{F_1} \neq \beta_{F_2} \neq \beta_{F_3}$, $c_1 = c_2 = c_3 \geq 1$ and $\rho_{PS} = \rho_{F_1} = \rho_{F_2} = \rho_{F_3}$, the parameters in this last equation are all denoted by ρ in the table.

p_1	p_2	p_3	β_{PS}	β_{F_1}	β_{F_2}	β_{F_3}	c_i	ρ	sim	appr	$\Delta\%$	simple	$\Delta\%$
0.1	0.2	0.3	0.08	0.80	0.40	0.27	1	0.20	1.97	1.99	0.87	1.38	-29.82
0.1	0.2	0.3	0.2	4.00	2.00	1.33	2	0.50	51.39	52.73	2.61	35.62	-30.68
0.1	0.2	0.4	0.24	7.20	3.60	1.80	3	0.80	675.26	685.68	1.54	390.88	-42.11
0.1	0.2	0.4	0.06	2.40	1.20	0.60	4	0.20	13.85	13.85	0.01	9.37	-32.34
0.1	0.3	0.4	0.1	5.00	1.67	1.25	5	0.50	122.15	126.17	3.29	71.35	-41.59
0.1	0.3	0.4	0.16	9.60	3.20	2.40	6	0.80	949.43	987.05	3.96	494.13	-47.96
0.1	0.3	0.5	0.02	1.40	0.47	0.28	7	0.20	25.66	25.84	0.69	12.77	-50.24
0.2	0.3	0.4	0.05	2.00	1.33	1.00	8	0.50	208.88	213.75	2.33	95.02	-54.51

Table 3.3 Variance of the total sojourn time for an asymmetric network with three different multi-server FCFS nodes with equal loads: approximations versus simulations

The results presented in table 3.3 show that even for an asymmetric network our approximation is still very accurate.

In table 3.4 we show the results for three FCFS nodes, with the same number of servers at each of the nodes. The probability of feedback is 0.9 for all cases. Now the probabilities of a visit to each of the FCFS nodes are equal, but the mean service times at the FCFS nodes are not equal, and thus the loads of the three FCFS nodes are not. For every run in table 3.4 it holds that $\lambda = 1$, $p_1 = p_2 = p_3 = 0.3$, $\beta_{F_1} \neq \beta_{F_2} \neq \beta_{F_3}$, $c_1 = c_2 = c_3 \geq 1$ and $\rho_{F_1} \neq \rho_{F_2} \neq \rho_{F_3}$.

β_{F_1}	β_{F_2}	β_{F_3}	c_i	ρ_{PS}	ρ_{F_1}	ρ_{F_2}	ρ_{F_3}	sim	appr	$\Delta\%$	simple	$\Delta\%$
0.75	0.50	0.25	3	0.50	0.75	0.50	0.25	79.60	80.62	1.28	36.54	-54.10
1.25	1.00	0.75	5	0.75	0.75	0.60	0.45	233.70	247.77	6.02	100.40	-57.04
2.4	1.33	0.27	8	0.75	0.90	0.50	0.10	610.90	655.33	7.27	329.00	-46.14
3.00	2.00	1.00	10	0.75	0.90	0.60	0.30	1025.38	1102.00	7.47	492.62	-51.96

Table 3.4 Variance of the total sojourn time for a network with three different multi-server FCFS nodes: approximations versus simulations

The results presented in table 3.4 show that for these asymmetric cases the approximation for the variance of the total sojourn time is still quite close to the simulation result. Nevertheless the results in table 3.4 seems to imply that either the approximation is less accurate for large values for the variance or the approximation is less accurate for a large number of servers at the FCFS nodes. Both possible reasons do not directly follow from tables 3.2 and 3.3.

Finally, we consider even more asymmetric network scenarios. In table 3.5 we present the results for three asymmetric FCFS nodes. We show different kinds of asymmetry in the network, as can be seen in the table below.

p_1	p_2	p_3	β_{F_1}	β_{F_2}	β_{F_3}	c_1	c_2	c_3	ρ_{PS}	ρ_{F_1}	ρ_{F_2}	ρ_{F_3}	sim	appr	$\Delta\%$
0.1	0.2	0.3	3.00	2.00	1.00	1	2	3	0.50	0.75	0.50	0.25	135.49	131.78	-2.74
0.1	0.2	0.4	2.25	2.70	1.69	1	3	5	0.75	0.75	0.60	0.45	211.61	216.96	2.53
0.1	0.3	0.4	1.80	0.40	0.15	1	8	9	0.75	0.90	0.08	0.03	638.54	642.47	0.61
0.1	0.3	0.4	1.80	3.20	1.35	1	8	9	0.75	0.90	0.60	0.30	904.46	909.08	0.51
0.2	0.3	0.4	1.00	1.67	2.00	10	10	10	0.50	0.20	0.50	0.80	386.12	429.41	11.21
0.3	0.3	0.3	0.40	0.67	0.53	6	4	2	0.80	0.20	0.50	0.80	194.92	199.34	2.26

Table 3.5 Variance of the total sojourn time for an asymmetric network with three multi-server FCFS nodes: approximations versus simulations

The results presented in table 3.5 show that, in most considered cases, our approximation is even for completely asymmetric networks very accurate. Unfortunately there is no space in the table to compare our approximation with the simple one; the error is between –50% and –25%. As can be seen, the difference between the simulated variance and the approximated variance is rather large for the case in table 3.5 with a high number of servers at all three FCFS nodes.

Remark on Bernoulli feedback

The developed accurate approximation for the variance of the total sojourn time can be applied to a Bernoulli feedback network with a single PS node and multiple multi-server FCFS nodes. Note that the feedback mechanism in the example queueing network described in chapter 1 is deterministic. So the approximation obtained in section 3.3.2 cannot directly be used for the decision what the contents of SLA's in the LWS-example will be. However, a network with Bernoulli feedback is a good model if many different applications use the same databases.

The approximation for the variance of the sojourn time in the Bernoulli feedback network can be used as an upperbound for the variance of the sojourn time in a network with deterministic feedback and similar loads, although simulation results showed that it is a very rough upperbound.

We consider the example queueing network model described in chapter 1. The loads in a deterministic network are defined as follows

$$\rho_{PS} = (N+1)\lambda\beta_{PS}, \tag{3.31}$$

$$\rho_{F_k} = N_k \lambda \beta_{F_k}, \; k = 1,\ldots,M. \tag{3.32}$$

N is the number of feedbacks to the PS node and N_k is the number of visits to the k^{th} FCFS node. In the example queueing network described in chapter 1 we assumed that the three FCFS nodes in the network were all visited once, and the PS node four times. So $N+1 = 4$ and $N_k = 1$ for $k = 1,2,3$. To approximate the model with deterministic feedback by a model with Bernoulli feedback, we take p and p_k such that $\frac{1}{1-p} = N+1$ and $\frac{p_k}{1-p} = N_k, \; k = 1,\ldots,M.$

Hence in our example $p = 3/4$ and $p_k = 1/4$ for $p = 1,2,3$. Then we can apply equations (3.11) and (3.13) to get the following expressions for the mean sojourn times for the i^{th} visit to a node

$$E\{S_{PS}{}^{i}\} = 4\frac{\beta_{PS}}{1-\rho_{PS}}, \; i = 1,2,\ldots, \tag{3.33}$$

$$E\{S_{F_k}{}^{i}\} = \frac{\beta_{F_k}}{1-\rho_{F_k}}, \; i = 1,2,\ldots, \; k = 1,.2,3. \tag{3.34}$$

Indeed, these expressions are exact for the product-form network under consideration.

If we apply (3.29) by taking $p = 3/4$ and $p_k = 1/4$ for $p = 1,2,3$, then we get a rough upperbound for the variance of the total sojourn time in the network with deterministic feedback. We can also take a very simple expression of the variance of the total sojourn time in a network with feedback as follows

$$Var_{lw}\{S\} \approx \frac{1}{1-p}\frac{2+\rho_{PS}}{2-\rho_{PS}}\left(\frac{\beta_{PS}}{1-\rho_{PS}}\right)^2 + \sum_{k=1}^{M}\frac{p_k}{1-p}\left(\beta_{Fk}{}^2 + \frac{p_w(2-p_w)\beta_{Fk}{}^2}{c_k{}^2(1-\rho_{Fk})^2}\right). \tag{3.35}$$

Then we get a rough lowerbound for the variance of the total sojourn time in the network with deterministic feedback. In the table below can be seen that this lowerbound is much closer to the simulated value than the suggested upperbound is.

β_{F_1}	β_{F_1}	β_{F_2}	β_{F_3}	c_1	c_2	c_3	ρ_{PS}	ρ_{F_1}	ρ_{F_2}	ρ_{F_3}	*Sim*	*Upper bound*	*Lower bound*
0,10	0,50	0,40	0,30	1	1	1	0,40	0,50	0,40	0,30	1,99	12,24	1,79
0,05	0,20	0,20	0,20	1	1	1	0,20	0,20	0,20	0,20	0,22	1,44	0,21
0,13	1,00	1,00	1,00	2	2	2	0,50	0,50	0,50	0,50	5,45	38,10	5,08
0,20	1,50	1,50	1,50	2	2	2	0,80	0,75	0,75	0,75	48,01	324,86	39,64
0,05	1,00	1,00	1,00	4	4	4	0,20	0,25	0,25	0,25	3,04	16,85	3,03
0,13	2,00	2,00	2,00	4	4	4	0,50	0,50	0,50	0,50	13,59	87,92	13,37
0,20	3,00	3,00	3,00	4	4	4	0,80	0,75	0,75	0,75	65,56	481,84	56,84
0,20	1,50	1,00	0,75	2	4	6	0,80	0,75	0,25	0,13	29,18	141,52	21,00
0,20	1,00	2,00	3,00	2	3	4	0,80	0,50	0,67	0,75	42,19	258,26	33,49
0,20	7,00	7,00	7,00	10	9	8	0,80	0,70	0,78	0,88	213,56	1625,84	208,58
0,05	2,00	2,00	2,00	10	10	10	0,20	0,20	0,20	0,20	12,02	63,08	12,02

Table 3.6 Numerical results for a network with a PS node and three multi-server FCFS nodes and deterministic feedback

By extending expression (3.35), we think that it is possible to obtain a good approximation for the variance of the total sojourn time in a network with deterministic routing. This is an interesting subject for further research.

Chapter 4

A feedforward queueing network with general service times

4.1 Introduction

In this chapter we study feedforward networks with a PS node and a single-server FCFS node, with generally distributed interarrival times and generally distributed service times at both nodes. In feedforward networks customers never return to a queue they have once left, hence feedforward networks are acyclic networks. In the next section we describe the models considered. In section 4.3 we describe the analysis. We start with sojourn time approximations for both GI/G/1-PS and GI/G/1-FCFS queues. Then we give the relations between the second moment of the arrival and the departure process for both the PS and the FCFS node. Next we describe the processes of splitting and superposition. Finally, we give some numerical results for the networks considered.

4.2 Model description

First we introduce the notation used in this chapter.

Model input parameters

λ	External arrival rate;
β_{PS}	Mean service time at the PS node;
β_F	Mean service time at the FCFS node;
$c_{b_F}^{\ 2}$	Squared coefficient of variation of the service time distribution at the FCFS node;
$c_{a_{PS}}^{\ 2}$	Squared coefficient of variation of the interarrival time distribution at the PS node;
$c_{a_F}^{\ 2}$	Squared coefficient of variation of the interarrival time distribution at the FCFS node;
$c_{d_{PS}}^{\ 2}$	Squared coefficient of variation of the interdeparture time distribution at the PS node;
$c_{d_F}^{\ 2}$	Squared coefficient of variation of the interdeparture time distribution at the FCFS node;
$\alpha(\zeta)$	Laplace-Stieltjes transform (LST) of the distribution function of the interarrival times;
ρ_{PS}	Load at the PS node;
ρ_F	Load at the FCFS node.

Random variables

S_{PS}	Total sojourn time in the PS node;
B_{PS}	Total service time in the PS node;
S_F	Total sojourn time in the FCFS node;
W_F	Total waiting time in the FCFS node;
B_F	Total service time in the FCFS node;
S	Total sojourn time in the network.

The squared coefficient of variation (SCV) is a measure of the variability. The SCV of a random variable rv is defined by:

$$c_{rv}^{\ 2} = \frac{\sigma^2(rv)}{E^2(rv)}. \tag{4.1}$$

Here $\sigma^2(rv)$ denotes the variance of a random variable rv and $E(rv)$ denotes the mean of this variable.

Simple network

We first consider a tandem network, i.e. a network in which all customers pass the stations once in the same order, that consists of a PS node and a single-server FCFS node, see figure 4.1. All customers enter the network at the PS node, receive service at the PS node, go the FCFS node, receive service at the FCFS node, and finally leave the network. The interarrival time as well as the service times at both nodes are generally distributed.

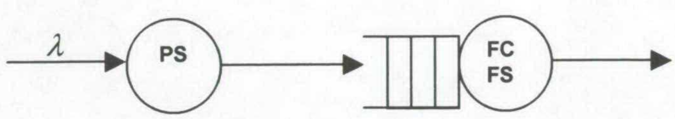


Figure 4.1 Tandem network with one PS node and one single-server FCFS node

The mean arrival rate at the PS node in figure 4.1 is equal to λ and the mean service time at the PS node is equal to β_{PS}. Hence it can be seen that the load at the PS node is given by

$$\rho_{PS} = \lambda\beta_{PS}. \tag{4.2}$$

The arrival rate at the FCFS node in figure 4.1 is also equal to λ and the mean service time at the FCFS node is equal to β_F. Hence it can be seen that the load at the FCFS node is given by

$$\rho_F = \lambda\beta_F. \tag{4.3}$$

To ensure stability of the network, we assume that the load at each node is smaller than one.

Extended network

We extend the network described above by adding another arrival stream and by adding the possibility that a customer leaves the network after visiting the PS node. This network is given in figure 4.2. The network consists of a single PS node and a single-server FCFS node as well. All customers enter the network at the PS node either from arrival process 1 with arrival rate λ_1 or from arrival process 2 with arrival rate λ_2. All customers receive service at the PS node. After service completion at the PS node, each customer goes to the FCFS node with probability p or leaves the network with probability 1-p. If a customer goes to the FCFS node, he leaves the network after service completion at that node.

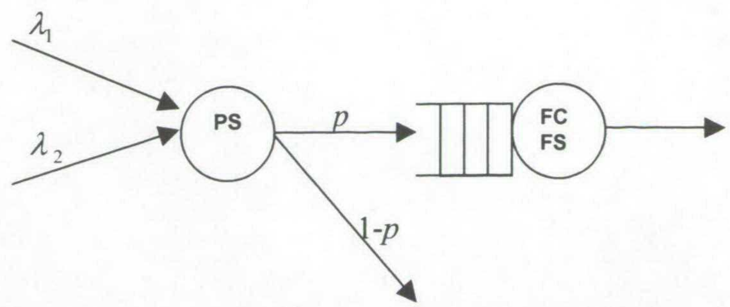


Figure 4.2 Network with two arrival steams with arrival rates λ_1 and λ_2, with probability p *customers departing from the PS node go to the single-server FCFS node*

The mean arrival rate at the PS node in figure 4.2 is equal to $\lambda_1 + \lambda_2$ and the mean service time at the PS node is equal to β_{PS}. Now it can be seen that the load at the PS node is given by

$$\rho_{PS} = (\lambda_1 + \lambda_2)\beta_{PS}. \tag{4.4}$$

The arrival rate at the FCFS node in figure 4.2 is equal to $p(\lambda_1 + \lambda_2)$ and the mean service time at the FCFS node is equal to β_F. Now it can be seen that the load at the FCFS node is given by

$$\rho_F = p(\lambda_1 + \lambda_2)\beta_F. \tag{4.5}$$

To ensure stability of the network, we assume that the load at each node is smaller than one.

4.3 Analysis

4.3.1 Sojourn time approximations

The total sojourn time in a network is equal to the sum of the sojourn times in the nodes of the network:

$$S = S_{PS} + S_F. \tag{4.6}$$

Generally distributed interarrival times lead to non-product-form networks. Hence both networks under consideration do not have a closed-form expression for the joint steady state distribution of the number of customers in the nodes, so that we cannot apply Little's law to obtain the expression for the mean sojourn times in the nodes. In the absence of exact results on the mean total sojourn time, we propose approximations for the mean sojourn times in both nodes.

GI/G/1-FCFS queue

For an approximation of the mean sojourn time in the FCFS node in the networks described in the previous section, we use an approximation for a GI/G/1 system. For a GI/G/1-FCFS server in a network, especially Whitt's QNA paper [28] contains much useful results, see section 2.2.4. Whitt gives the following approximation for the mean waiting time in the GI/G/1 queue

$$E\{W_F\} \approx \frac{\beta_F \rho_F \left({c_{a_F}}^2 + {c_{b_F}}^2\right) g}{2(1-\rho_F)}, \tag{4.7}$$

where $g = \begin{cases} \exp\left[-\dfrac{2(1-\rho_F)}{3\rho_F}\dfrac{(1-c_{a_F}{}^2)^2}{c_{a_F}{}^2+c_{b_F}{}^2}\right], & c_{a_F}{}^2 < 1 \\ 1, & c_{a_F}{}^2 \geq 1. \end{cases}$ (4.8)

When $c_{a_F}{}^2 < 1$, (4.7) is the Kraemer and Langenbach-Belz approximation [18], which is known to perform well [28]. When $c_{a_F}{}^2 > 1$, the original Kraemer and Langenbach-Belz refinement does not seem to help, so it is not used. Note that (4.7) is exact for the M/G/1-FCFS queue having $c_{a_F}{}^2 = 1$.

We assume that the service time distribution is given, so β_F and $c_{b_F}{}^2$ are known. If we also assume that the interarrival time distribution is given, then λ and $c_{a_F}{}^2$ are known too. Then we can calculate the load ρ_F at the FCFS node. Subsequently, we can substitute everything in (4.7) and get an approximation for the mean waiting time. Note that the sojourn time in an FCFS node is equal to the waiting time plus the service time. Hence we know that

$$E\{S_F\} = E\{W_F\} + E\{B_F\} \approx \frac{\beta_F \rho_F \left(c_{a_F}{}^2 + c_{b_F}{}^2\right) g}{2(1-\rho_F)} + \beta_F, \tag{4.9}$$

where g is defined as in (4.8).

GI/G/1-PS queue

For an approximation of the mean sojourn time in the PS node in the networks described in the previous section, we take into account an isolated GI/G/1-PS node. Sengupta [24] proposes an approximation for the sojourn time distribution for the GI/G/1-PS node. As input for the approximation of the mean sojourn time we need the LST of the interarrival time distribution and the mean service time at the PS node. We assume that the mean service time at the PS node, β_{PS}, is known. Assume that we also know the probability distribution of the interarrival times at the PS node. Then the LST of the interarrival times $\alpha(\zeta)$ is also known.

Now let η denote the smallest positive root of the equation

$$\eta = \alpha\left(\frac{1}{\beta_{PS}}(1-\eta)\right). \tag{4.10}$$

Sengupta concludes that

$$S_{PS} \stackrel{d}{\approx} \frac{B_{PS} X}{1-\eta}. \tag{4.11}$$

B_{PS} denotes the service time at the PS node. X denotes a random variable, independent of B_{PS}, with expected value equal to one. Hence it follows that

$$E\{S_{PS}\} = E\left\{\frac{B_{PS} X}{1-\eta}\right\} = \frac{E\{B_{PS}\}E\{X\}}{1-\eta} = \frac{\beta_{PS}}{1-\eta}, \tag{4.12}$$

where η is the smallest positive root of equation (4.10).

Note that (4.12) is exact for the M/G/1-PS queue and the GI/M/1-PS queues, see Ott [22], Jagerman and Sengupta [16] and Ramaswami [23]. For the M/G/1-PS case this can easily be seen since $\eta = \rho_{PS}$ if the arrival process is a Poisson process. Sengupta's approximation for the mean sojourn time gives the same result for a GI/G/1-PS queue and a GI/M/1-PS queue, since it is only based on the mean service time and not on the entire service time distribution.

4.3.2 Departure processes at the nodes

In the previous subsection we showed how we can approximate the mean sojourn times in both nodes of the networks described in section 4.2. Therefore we have to know (an approximation of) the SCV's of the arrival process and the service process at both nodes. We assume that the service time distribution is known. For the tandem network given in figure 4.1, it can easily be seen that the arrival process to the FCFS node is equal to the departure process from the PS node. As far as we know, there is no expression for the SCV of the departure process from the PS node known yet.

For three different probability distributions of the service time in the PS node ($c_{b_{PS}}^2$ smaller than, equal to and larger than 1) we simulated an isolated PS node. For different SCV's of the arrival process we measured the SCV's of the departure process. We did this for different loads. For every probability distribution and for every load considered the relation between the both SCV's was very close to

$$c_{d_{PS}}^2 \approx \rho_{PS}^2 + (1-\rho_{PS}^2)c_{a_{PS}}^2 . \tag{4.13}$$

Equation (4.13) is exact for Poisson arrival processes, where $c_{a_{PS}}^2 = 1$. Then (4.13) implies $c_{d_{PS}}^2 = 1$. If $\rho_{PS} = 0$, (4.13) is exact as well, because then $c_{d_{PS}}^2 = c_{a_{PS}}^2$. If $\rho_{PS} = 1$, (4.13) is exact only if the service times are exponentially distributed, because then $c_{d_{PS}}^2 = c_{b_{PS}}^2 = 1$.

.

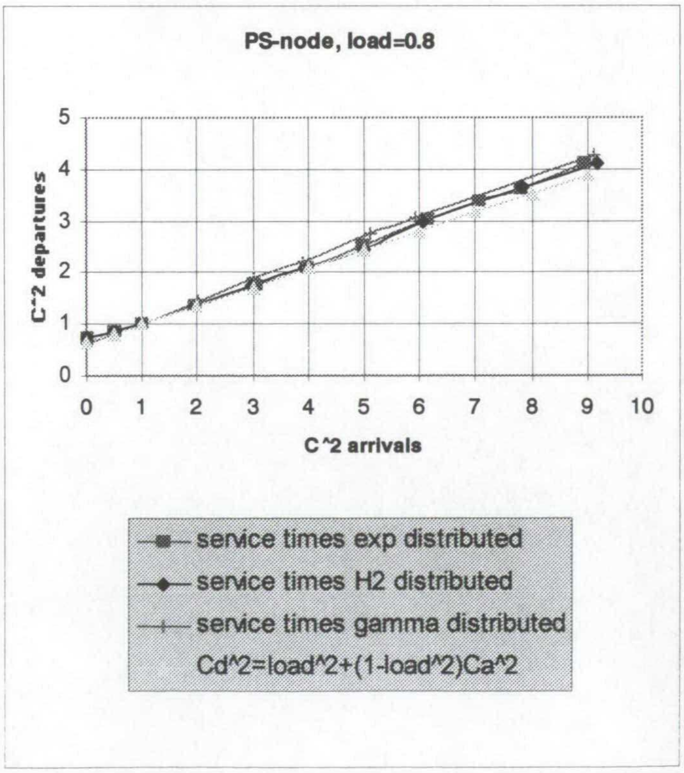


Figure 4.3 Relation between SCV's for arrival and departure process in a PS node

In figure 4.3 above relation (4.13) as well as the simulation results for different SCV's of the arrival process are given for a PS node with a load equal to 0.8. In appendix C some pictures of this relationship can be found.

Whitt obtained an approximate formula for the SCV of the departure process from an FCFS node

$$c_{d_F}^{\ 2} \approx \rho_F^{\ 2} c_{b_F}^{\ 2} + (1-\rho_F^{\ 2}) c_{a_F}^{\ 2}. \tag{4.14}$$

Note that the approximate formulas for the SCV of the departure process from a PS node (4.13) and from a FCFS node (4.14) are very similar, but there is a difference. The departure process from the FCFS node depends on the service time variability in that node. Apparently, that does not hold for the PS node. The SCV of the departure process at the PS node just depends on the load in that node and the arrival process to that node.

We will use (4.13) to determine an approximation for the SCV of the departure process at a PS node, and thus to determine the SCV of the arrival process at the FCFS node.

4.3.3 Flow through a network

If we extend the tandem network given in figure 4.1 to the network in figure 4.2, then we need equations for the SCV's in case of splitting or merging in the network.

Splitting

Whitt [28] gives a formula for splitting. He remarks that no further approximation is needed for splitting, because a renewal process that is split by independent probabilities is again a renewal process. Note that the process being split is typically not a renewal process! A relation for the SCV's is easy to obtain, because the renewal-interval distribution in the split stream is a geometrically distributed random sum of the original renewal intervals. Whitt's formula is applicable for a stream being split into k streams. In general, if a stream with a parameter c^2 is split into k streams, with each being selected independently according to probabilities p_i, i = 1, 2, ..., k, then the i^{th} process obtained from the splitting has SCV c_i^2 given by:

$$c_i^{\ 2} \approx p_i c^2 + 1 - p_i. \tag{4.15}$$

Superposition

Whitt [28] also gives formulas for superposition, i.e. the merging of two processes. Whitt remarks that neither the asymptotic method nor the stationary-interval method alone works well over a wide range of cases. We do not go into details about these methods, more information about these methods can be found in [28]. Whitt suggests taking v times the c^2 obtained by the asymptotic method and 1-v times the SCV obtained by the stationary-interval method. However, Whitt refers to Albin [1] [2], who found that a convex combination of the SCV obtained by the asymptotic method and the exponential SCV of 1 worked almost as well. As for superposition the stationary-interval method is nonlinear, it presents difficulties. For these two reasons, instead of the non-linear SCV obtained by the stationary-interval method Whitt simply uses the value one. As a conclusion Whitt comes up with the following formula for the superposition SCV c_j^2 as a function of component SCV's c_i^2 and the arrival rates λ_i :

$$c_j^{\ 2} \approx v_j \sum_i \left(\frac{\lambda_i}{\sum_k \lambda_k} \right) c_i^{\ 2} + 1 - v_j, \tag{4.16}$$

where $v_j = \left[1 + 4(1-\rho_j)^2\left(\left[\sum_i p_{ij}^2\right]^{-1} - 1\right)\right]^{-1}$. (4.17)

Whitt [28] suggests v_j as given in (4.17), this is based on extensive simulations by Albin as well as on theoretical results by himself. We simply take the v_j Whitt uses in [28]. In [29] he shows that this formula gives quite accurate results. The parameter p_{ij} is defined as the proportion of arrivals to node j that came from node i, $i \geq 0$.

4.3.4 Approximation for mean total sojourn time

For obtaining the total mean sojourn time in a network, we can simply sum the mean sojourn times in the nodes in the network is mentioned in (4.6). Approximations for these mean sojourn times are given in section 4.3.1

$$E\{S\} = E\{S_{PS}\} + E\{S_F\} = \frac{\beta_{PS}}{1-\eta} + \frac{\beta_F \rho_F \left(c_{a_F}^{\ 2} + c_{b_F}^{\ 2}\right)g}{2(1-\rho_F)} + \beta_F. \tag{4.18}$$

As input for these approximations we need the SCV's of the arrival processes to the nodes, therefore we need the expressions given in section 4.3.2 and 4.3.3.

Simple network

For the simple network, given in figure 4.1, we know the external arrival process. The SCV of the arrival process at the PS node is equal to the SCV of the external interarrival times. By using (4.13) we approximate the SCV of the interdeparture times from the PS node. The SCV of the interarrival times at the FCFS node is equal to the SCV of the interdeparture times from the PS node.

Hence the mean sojourn time in the network given in figure 4.1 can be approximated as follows

$$E\{S\} = \frac{\beta_{PS}}{1-\eta} + \frac{\beta_F \rho_F \left(\rho_{PS}^{\ 2} + (1-\rho_{PS}^{\ 2})c_{a_{PS}}^{\ 2} + c_{b_F}^{\ 2}\right)g}{2(1-\rho_F)} + \beta_F. \tag{4.19}$$

where η is the smallest positive root of equation (4.10) and g is defined as in (4.8).

Extended network

For the extended network, given in figure 4.2, we know both external arrival process. The arrival process at the PS node is equal to the merging of both external arrival processes, therefore we need (4.16). By using (4.13) we approximate the SCV of the interdeparture times from the PS node. The arrival process at the FCFS node is equal to a geometrical part of the departure process from the PS node, therefore we use (4.15).
Hence the mean total sojourn time in the network given in figure 4.2 can be approximated as follows

$$\begin{aligned} E\{S\} = \frac{\beta_{PS}}{1-\eta} + \frac{\beta_F \rho_F g}{2(1-\rho_F)}\Bigg(p\Big(\rho_{PS}^{\ 2} + (1-\rho_{PS}^{\ 2})\Big(\left[1-2(1-\rho_{PS})^2\right]^{-1}\left(\frac{\lambda_1 c_1^{\ 2}}{\lambda_1+\lambda_2} + \frac{\lambda_2 c_2^{\ 2}}{\lambda_1+\lambda_2}\right) \\ + 1 - \left[1-2(1-\rho_{PS})^2\right]^{-1}\Big)\Big) + (1-p) + c_{b_F}^{\ 2}\Bigg) + \beta_F, \end{aligned} \tag{4.20}$$

where η is the smallest positive root of equation (4.10), with $\alpha(\zeta)$ the LST of the total arrival process at the PS node. The parameter g is defined in (4.8).

4.4 Numerical results

4.4.1 Remarks about Sengupta's approximation

By using Sengupta's approximation (4.12) for the mean sojourn time in an isolated PS node, we discovered that his method does not work for PS nodes with big SCV for the arrival process. In tables 4.1 and 4.2 we compare simulation results with Sengupta's approximation for the mean sojourn time in an isolated PS-node. In the tables we show some remarkable results.

For every case the mean service time is taken equal to one. Thus the load in a single server PS node, see section 4.2, is equal to the interarrival rate. The symbol $\Delta\%$ in the tables below denotes the percentage error, which is defined in formula (3.30).

ρ_{PS}	$c_{a_{PS}}^{2}$	simulation	approximation	$\Delta\%$
0.5	2	2.36	2.45	3.83
0.5	4	2.92	3.16	8.24
0.5	6	3.39	3.74	10.45
0.8	2	6.18	6.97	12.76
0.8	4	8.78	10.82	23.21
0.8	6	11.09	14.62	31.80

Table 4.1 Sengupta's approximation for mean sojourn times compared to simulation results, $c_{b_{PS}}^{2}$ *= 2.125.*

ρ_{PS}	$c_{a_{PS}}^{2}$	simulation	approximation	$\Delta\%$
0.5	2	2.50	2.45	-2.19
0.5	4	3.29	3.16	-3.83
0.5	6	4.19	3.74	-10.79
0.8	2	7.15	6.97	-2.50
0.8	4	11.89	10.82	-9.00
0.8	6	17.87	14.62	-18.19

Table 4.2 Sengupta's approximation for mean sojourn times compared to simulation results, $c_{b_{PS}}^{2}$ *= 0.5.*

In table 4.1 the service times are hyperexponentially distributed, so the SCV for the service times is larger than one. We vary the SCV for the interarrival times. In table 4.2 the service times have a gamma distribution with the SCV smaller than one. Again we vary the SCV for the interarrival times.

Note that Sengupta's approximation gives the same results for mean sojourn time in both tables, although the SCV's for the service process are different. For a SCV for the service process bigger than one, the approximation overestimates the mean sojourn time in the PS node. For a SCV for the service process smaller than one, the approximation underestimates the mean sojourn time in the PS node. Sengupta notes that his mean sojourn time approximation is exact for the M/G/1 and the GI/M/1 cases. Cohen remarks in the second edition of [10] that the mean sojourn time in a M/G/1-PS node only depends on the first moment of the service time distribution, not on the distribution itself. Our simulation results show that this is not the case for the GI/G/1-PS queue, however it seems that Sengupta assumed this.

To verify if there was no mistake in our simulation model we reconstructed some of Sengupta's data. For the sojourn time distribution of H_2/M/1-PS and H_2/H_2/1-PS queues some tail probabilities are given in table 4.3.

As shown in table 4.3, our simulation experiments gave almost exactly the same results for some tail probabilities as his experiments did. So we conclude that our simulation model is valid. Sengupta either did not test his method for large values for the squared coefficients or he was content with a rather small accuracy for big values of the SCV. In practice, the SCV's of an arrival process will probably not be very big. As can be seen in the tables in [24], Sengupta's method gives appropriate results for tail probabilities in case of SCV's of the interarrival and service process smaller than or equal to two.

λ	$c_{a_{PS}}^2$	$c_{b_{PS}}^2$	x	$P\{sojourn\,time \le x\}$ Sengupta's approximation	Sengupta's simulation	Our simulation	$\Delta\%$
0.8	2.0	1.0	1.5	0.320	0.318	0.318	0.63
0.8	2.0	1.0	3.0	0.483	0.485	0.485	-0.41
0.8	2.0	1.0	4.5	0.587	0.594	0.594	-1.18
0.8	2.0	2.0	1.5	0.370	0.382	0.384	-3.14
0.8	2.0	2.0	3.0	0.534	0.551	0.551	-3.09

Table 4.3 Some tail probabilities for the sojourn time in a PS-node with general distributed interarrival times

4.4.2 Results for simple network

For the network described in section 4.2 and given in figure 4.1, we compared the approximations for the SCV's of the departure processes from both nodes with simulation results. We also compared approximations for the mean sojourn times in both nodes with simulation results.

λ	$c_{a_{PS}}^2$	ρ_{PS}	$c_{b_{PS}}^2$	ρ_F	$c_{b_F}^2$	$c_{d_{PS}}^2 = c_{a_F}^2$		$c_{d_F}^2$		$E\{S_{PS}\}$		$E\{S_F\}$	
						sim	*appr*	*sim*	*appr*	*sim*	*appr*	*sim*	*appr*
0.40	1.38	0.32	2.13	0.50	1.67	1.33	1.34	1.40	1.42	1.24	1.24	3.15	3.13
0.50	2.13	0.40	2.13	0.63	1.67	1.90	1.95	1.84	1.84	1.54	1.59	5.11	5.01
0.30	1.08	0.30	2.13	0.45	2.92	1.07	1.08	1.45	1.45	1.45	1.45	3.98	3.95
0.40	1.38	0.40	2.13	0.60	2.92	1.30	1.32	1.88	1.90	1.77	1.79	6.30	6.27
0.50	2.13	0.50	2.13	0.75	2.92	1.84	1.84	2.39	2.45	2.38	2.50	12.38	12.22
0.60	4.56	0.60	2.13	0.90	2.92	3.39	3.28	3.08	2.99	4.16	4.75	47.79	43.33

Table 4.4 Approximations compared to simulation results for a simple network

As we expected, the approximations are very accurate for the simple network. For the last case, the approximations for the mean sojourn times are not that good. For the mean sojourn time in the PS node, we think that it is due to Sengupta's approximation, see the previous subsection for remarks about that approximation. In the pictures in appendix C, it can be seen that (4.13) becomes worse for large SCV of the arrival process at the PS node as well. In the table it can be seen that Whitt's approximation for the mean sojourn time in an FCFS node is not very accurate for large values of SCV's too.

We compared the approximations for the mean total sojourn times given in the previous section with the exact expression for the mean total sojourn time in case of M/M/1 nodes

$$E\{S\} = \frac{\beta_{PS}}{1-\rho_{PS}} + \frac{\beta_F}{1-\rho_F}. \tag{4.21}$$

The approximations developed in this chapter are much more accurate than (4.21) in all considered cases.

4.4.3 Results for extended network

For the extended network described in section 4.2 and given in figure 4.2, we compared the approximations for the SCV's of the arrival processes to and departure processes from both nodes with simulation results. We also compared approximations for the mean sojourn times in both nodes with simulation results. In the tables below we give the characteristics and the results respectively for a number of parameter combinations.

Nr.	λ_1	$c_{a_1}^2$	λ_2	$c_{a_2}^2$	ρ_{PS}	$c_{b_{PS}}^2$	ρ_F	$c_{b_F}^2$	p
1	1	1	0.50	1.38	0.75	2.13	0.60	1.67	0.8
2	1	1	0.50	1.38	0.30	2.13	0.48	1.67	0.8
3	1	1	1.00	1.38	0.80	2.13	0.80	1.67	0.8
4	1	1	1.50	1.38	0.75	2.13	0.80	1.67	0.8
5	1	1	2.00	1.38	0.60	2.13	0.96	1.67	0.8
6	1	1	0.50	1.67	0.75	2.92	0.38	2.13	0.5
7	1	1	0.50	1.67	0.75	2.92	0.60	2.13	0.8
8	1	1	1.00	1.67	0.80	2.92	0.80	2.13	0.8
9	1	1	1.50	1.67	0.75	2.92	0.80	2.13	0.8
10	1	1	1.00	2.13	0.80	2.13	0.80	1.67	0.8
11	1	1	0.50	2.92	0.75	2.92	0.75	2.13	0.5
12	1	1	0.50	4.56	0.75	2.92	0.60	2.13	0.8

Table 4.5 Characteristics of the extended network for different parameter combinations

Nr.	$c_{a_{PS}}^2$		$c_{d_{PS}}^2$		$c_{a_F}^2$		$c_{d_F}^2$		$E\{S_{PS}\}$		$E\{S_F\}$	
	sim	appr	sim	appr	sim	appr	sim	appr	sim	appr	sim	appr
1	1.04	1.15	1.01	1.06	1.03	1.05	1.26	1.27	2.06	2.03	1.51	1.50
2	1.03	7.35	1.03	6.78	1.03	5.62	1.18	4.71	0.29	0.29	0.90	0.90
3	1.09	1.21	1.05	1.07	1.04	1.06	1.46	1.45	2.09	2.07	3.27	3.19
4	1.12	1.26	1.04	1.11	1.04	1.09	1.43	1.46	1.28	1.25	2.68	2.57
5	1.16	1.37	1.10	1.24	1.09	1.19	1.61	1.63	0.53	0.52	13.68	13.58
6	1.05	1.25	1.03	1.11	1.03	1.06	1.19	1.21	2.06	2.04	0.98	0.97
7	1.07	1.25	1.03	1.11	1.03	1.09	1.46	1.46	2.10	2.04	1.70	1.68
8	1.13	1.36	1.05	1.13	1.04	1.10	1.72	1.76	2.12	2.11	3.80	3.66
9	1.18	1.46	1.09	1.20	1.08	1.16	1.73	1.78	1.29	1.29	3.02	2.96
10	1.21	1.61	1.10	1.22	1.09	1.18	1.48	1.49	2.26	2.16	3.43	3.22
11	1.11	1.73	1.06	1.32	1.07	1.16	1.68	1.70	2.23	2.09	6.05	5.73
12	1.16	2.35	1.10	1.59	1.10	1.47	1.49	1.71	2.35	2.11	1.83	1.69

Table 4.6 Results for the extended network, the row numbers correspond to the numbers in table 4.5

Although the approximations for the SCV's are not always accurate, the approximate SCV's lead to accurate approximations for the mean sojourn times in the nodes. Hence, not only for the simple network, but also for the extended network, the approximations for the mean sojourn times in the nodes are very accurate.

Again, we compared the approximations for the mean total sojourn times given in the previous section with the exact expression for the mean total sojourn time in case of exponentially distributed service times. In all cases, the approximations developed in this chapter are much more accurate than (4.21).

We think that the approximations for the SCV's of arrival and departure processes at and the mean sojourn times in PS nodes and single-server FCFS nodes are also applicable to networks consisting of more PS and single-server FCFS nodes with non-exponentially distributed interarrival and service times. However, we are not sure if the approximations still work well for networks in which overtaking can take place. Nevertheless, in the next chapter we will apply some of the results to a feedback network.

We think that the results presented in this chapter can easily be extend to approximations for variances of the total sojourn time in the networks described in section 4.2, this is an interesting topic for further research.

Chapter 5

A two-node feedback queueing network with general service times

5.1 Introduction

Boxma et al. [8] describe a method for approximating sojourn time distributions in open queueing networks. The authors restrict themselves to a two-node open queueing network with a PS node and a single-server FCFS node, we will described the model in the next section. In section 5.3 the approximations for the mean total sojourn time given in [8] are described and an improved approximation for the mean total sojourn time is given. In the next section we will describe the model. In section 5.4, we give some numerical results, which are used to validate the approximations for the mean total sojourn by Boxma et al. [8] and the improved approximation.

5.2 Model description

First we introduce the notation used in this chapter.

Model input parameters

λ External interarrival rate;

p Probability of feedback to the FCFS node after leaving the PS node;

β_{PS} Mean service time at the PS node;

β_F Mean service time at the FCFS node;

$\beta_F^{(2)}$ Second moment of the service time distribution in the FCFS node;

ρ_{PS} Load at the PS node;

ρ_F Load at the FCFS node.

Random variables

S Total sojourn time in the network;

S_{PS} Total sojourn time in the PS node;

S_F Total sojourn time in the FCFS node;

$S_{PS}^{(j)}$ Total sojourn time at the first *j* visits to the PS node;

$S_F^{(j)}$ Total sojourn time at the first *j* visits to the FCFS node;

σ_F Total sojourn time in the FCFS node short-circuited;

$\sigma_{PS}^{(j)}$ Total sojourn time in the PS node short-circuited at the first *j* visits to that node;

$\sigma_F^{(j)}$ Total sojourn time in the FCFS node short-circuited at the first *j* visits to that node.

Model

We consider a two-node open queueing network with a PS node and a single-server FCFS node. External customers arrive at the PS node according to a Poisson process with rate λ. A departing customer subsequently enters the FCFS node with probability p, and leaves the system with probability 1-p. Upon departure from the FCFS node, a customer always returns to the PS node. The service times at all visits to both nodes are independent random variables with means β_{PS} and β_F, at the PS and FCFS node respectively.

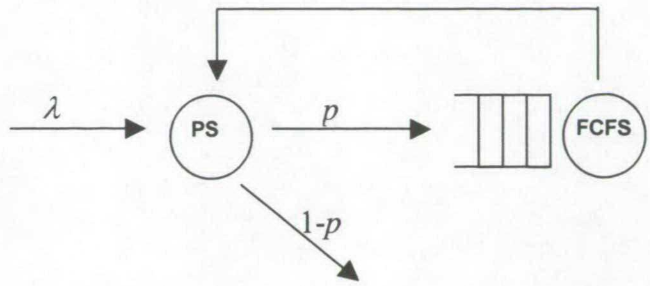


Figure 5.1 Feedback network considered with a PS node and a single-server FCFS node

The external arrival rate is equal to λ. Beside the external arrivals at the PS node, there are also internal arrivals, i.e. arrivals that are fed back with probability p. So the mean arrival rate at the PS node is

$$\lambda_{PS} = \lambda + p\lambda + p^2\lambda + ... = \frac{\lambda}{1-p}. \quad (5.1)$$

The mean service time at the PS node is equal to β_{PS}. Hence it can be seen that the total load at the PS node is given by

$$\rho_{PS} = \frac{\lambda\beta_{PS}}{1-p}. \quad (5.2)$$

The arrival rate at the FCFS node is equal to the probability that a customer goes to the FCFS node after visiting the PS node times the total arrival rate at the FCFS node

$$\lambda_F = \frac{p\lambda}{1-p}. \quad (5.3)$$

The mean service time at the FCFS node is equal to β_F. Hence it can be seen that the total load at the FCFS node is given by

$$\rho_F = \frac{p\lambda\beta_F}{1-p}. \quad (5.4)$$

5.3 Analysis

5.3.1 Distribution of the total sojourn time

It is hard to obtain exact results for the sojourn time distributions in queueing networks, if some form of overtaking of customers (or their effects) occurs, cf. [7] and section 2.2.2. In the model considered in this chapter, both the PS service discipline and feedback induce such overtaking. As mentioned before, an exact result for the mean sojourn times is not known in case of general service time distributions at the FCFS node. Hence we take recourse to approximations.

The approximation assumptions are:

- $S_{PS}^{(j+1)}$ and $S_F^{(j)}$ are independent, $j = 0,1,...$;
- $S_{PS}^{(j)}$ has the same distribution as $\sigma_{PS}^{(j)}$ and similarly $S_F^{(j)}$ has the same distribution as $\sigma_F^{(j)}$.

Now the Laplace-Stieltjes transform (LST) of the joint distribution of the total sojourn times in both nodes can be approximated as follows, for $\omega_1, \omega_2 \geq 0$

$$\begin{aligned} E[e^{-\omega_1 S_{PS} - \omega_2 S_F}] &= \sum_{k=0}^{\infty} (1-p)p^k E[e^{-\omega_1 S_{PS}^{(k+1)} - \omega_2 S_F^{(k)}}] \\ &\approx \sum_{k=0}^{\infty} (1-p)p^k E[e^{-\omega_1 \sigma_{PS}^{(k+1)}}] E[e^{-\omega_2 \sigma_F^{(k)}}]. \end{aligned} \tag{5.5}$$

Boxma et al. [8] refer to Doshi and Kaufman [12] for the LST for an M/G/1-FCFS queue with feedback and to Ott [22] for the LST for an M/G/1-PS queue, but they remark that these LST expressions are quite complicated. At this moment, we are satisfied with obtaining an expression for the mean of $S_{PS}^{(j)}$ and $S_F^{(j)}$, and hence of S_{PS} and S_F. Relatively easy expressions for the mean of these random variables can be taken from [22] and [25]. In the remainder of this section we will concentrate on the mean sojourn times. Hence we do not need to use the LST anymore. In the following subsections we will first describe the two approximations given in [8] and after that we give an improved approximation.

5.3.2 First approximation for mean total sojourn time

For approximating the mean total sojourn time, we need that

$$E\{S\} = \sum_{k=0}^{\infty} (1-p)p^k \left(E\{S_{PS}{}^{(k+1)}\} + E\{S_F{}^{(k)}\}\right) = E\{S_{PS}\} + E\{S_F\}. \tag{5.6}$$

Now we need just one approximation assumption:

- $S_{PS}^{(j)}$ has the same distribution as $\sigma_{PS}^{(j)}$ and similarly $S_F^{(j)}$ has the same distribution as $\sigma_F^{(j)}$.

For the PS node we know that

$$E_{appr1}\{S_{PS}\} = \sum_{k=0}^{\infty} (1-p)p^k E\left\{\sigma_{PS}{}^{(k+1)}\right\} = (1-p)\sum_{k=0}^{\infty} p^k (k+1)\frac{\beta_{PS}}{1-\rho_{PS}} = \frac{1}{1-p}\frac{\beta_{PS}}{1-\rho_{PS}}. \tag{5.7}$$

The second step follows from, see, e.g., section 3.2.1,

$$E\left\{\sigma_{PS}{}^{(k+1)}\right\} = \sum_{i=1}^{k+1} E\left\{\sigma_{PS}{}^{(i)}\right\} = \sum_{i=1}^{k+1} \frac{\beta_{PS}}{1-\rho_{PS}} = (k+1)\frac{\beta_{PS}}{1-\rho_{PS}}. \tag{5.8}$$

Note that (5.7) and (5.8) are exact for the network described in section 5.2 with exponentially distributed service times instead of generally distributed service times at the PS node.

In an earlier version of [8] we discovered a mistake in the approximation for the mean sojourn time in the FCFS node and a few other small mistakes. See appendix D for the old approximation with mistake and some explanation about the mistakes.

By using formula (35) of Takács [25] it follows for the FCFS node that

$$\begin{aligned} E_{appr1}\{S_F\} &= p\sum_{k=1}^{\infty}(1-p)p^{k-1}\mathrm{E}\{\sigma_F^{(k)}\} = p\left(\frac{p\lambda\beta_F^{(2)} + 2\beta_F(1-p\lambda\beta_F)}{2((1-p)-p\lambda\beta_F)}\right) \\ &= \frac{p}{1-p}\left(\frac{\beta_F}{1-\rho_F} + \frac{\lambda}{2}(\beta_F^{(2)} - 2\beta_F^{2})\frac{p}{1-\rho_F}\right). \end{aligned} \tag{5.9}$$

Note that for exponentially distributed service times at the FCFS node, $\beta_F^{(2)} - 2\beta_F^{2} = 0$. Then expression (5.9) is equal to the exact expression for the sojourn time in an M/M/1-FCFS node in a product-form network.

Combining (5.7) and (5.9) gives the following approximation for the total sojourn time in the given network

$$\begin{aligned} E_{appr1}\{S\} &= E_{appr1}\{S_{PS}\} + E_{appr1}\{S_F\} \\ &= \frac{1}{1-p}\frac{\beta_{PS}}{1-\rho_{PS}} + \frac{p}{1-p}\left(\frac{\beta_F}{1-\rho_F} + \frac{\lambda}{2}(\beta_F^{(2)} - 2\beta_F^{2})\frac{p}{1-\rho_F}\right). \end{aligned} \tag{5.10}$$

Numerical experiments, cf. [8], suggest that the used approximation assumption works well when one is mainly interested in approximating the total sojourn time, while one of the queues has much larger mean sojourn times at each visit than the other. Then the assumption should be quite accurate for that "bottleneck" queue. It is probably far from accurate for the other queue, but the contribution of the latter queue to the variance of the total sojourn time is probably rather small. So the given approximation for the mean total sojourn time works well in that specific case, but not in other cases.

5.3.3 Second approximation for mean total sojourn time

If the mean sojourn times at both queues are roughly equal, then the first approximation of the mean total sojourn time can be improved in the following way. Define a weight w as follows

$$w = \frac{E\{\sigma_{PS}^{(1)}\}}{E\{\sigma_{PS}^{(1)}\} + E\{\sigma_F^{(1)}\}}. \tag{5.11}$$

We know that

$$E\{\sigma_{PS}^{(1)}\} = \frac{\beta_{PS}}{1-\rho_{PS}}. \tag{5.12}$$

According to Van den Berg [4]

$$E\{\sigma_F^{(1)}\} = \frac{\beta_F}{1-\rho_F} + \frac{\lambda}{2}(\beta_F^{(2)} - 2\beta_F^{2})\frac{p}{1-p}\frac{1-p\rho_F}{1-\rho_F}. \tag{5.13}$$

Replacing the mean sojourn times at both nodes by weighted sums of mean sojourn times that correspond to the two extreme cases of short-circuiting (i.e., immediate feedback to the same queue) and independence of successive sojourn times of a customer at the same queue (i.e., feedback after a long time) leads to

$$E_{appr2}\{S_{PS}\} = \sum_{k=0}^{\infty}(1-p)p^k(wE\{\sigma_{PS}^{(k+1)}\} + (1-w)(E\{\sigma_{PS}^{(1)}\})^{k+1}) \tag{5.14}$$

and

$$E_{appr2}\{S_F\} = \sum_{k=0}^{\infty}(1-p)p^k((1-w)E\{\sigma_F^{(k)}\} + w(E\{\sigma_F^{(1)}\})^k). \tag{5.15}$$

Now we can motivate the choice of w as defined in (5.11). As can be seen, $w = 0$ corresponds to the case that the sojourn time for the first visit to the PS node is equal to zero, so one can conclude that the PS node "hardly" exists. In that case, the mean sojourn time in the PS node will be approximated by zero. The mean sojourn time in the FCFS node will be approximated by the mean sojourn time in an FCFS node with direct feedback, so as if the PS node does not exist. For $w = 1$ it is the other way around, then one can conclude that the FCFS node "hardly" exists. The mean sojourn time in the PS node will be approximated by the mean sojourn time in a PS node with direct feedback, so as if the FCFS node does not exist. The mean sojourn time in the FCFS node will be approximated by zero. So these two extreme cases are covered correctly by this choice of w. In general, the weight gives the ratio between the mean sojourn time of the first visit to the PS node and the mean total sojourn time of the first visit to both nodes. So if $w = 0.5$, this means that the first visit to the PS node by a customer takes the same time as the first visit to the FCFS node by that customer.

The short-circuiting case is exactly the same as the first approximation with the same assumption. For the other case, independence of successive sojourn times of a customer at the same queue, we need another assumption, namely:

- Customers arrive at both nodes according to a Poisson process.

This assumption is more or less the same as the first assumption in section 3.3.2. So we assume M/G/1 severs: exponentially distributed interarrival times and generally distributed service times at both nodes.

For the approximation of the mean sojourn time at the PS node (5.14), this method gives exactly the same result as before, i.e.,

$$\begin{aligned} \mathrm{E}_{appr2}\{S_{PS}\} &= \sum_{k=0}^{\infty}(1-p)p^k(w\mathrm{E}\{\sigma_{PS}{}^{(k+1)}\} + (1-w)(k+1)\mathrm{E}\{\sigma_{PS}{}^{(1)}\}) \\ &= (1-p)\sum_{k=0}^{\infty}p^k(w(k+1)\frac{\beta_{PS}}{1-\rho_{PS}} + (1-w)(k+1)\frac{\beta_{PS}}{1-\rho_{PS}}) = \frac{1}{1-p}\frac{\beta_{PS}}{1-\rho_{PS}}. \end{aligned} \tag{5.16}$$

The result for the approximation of the mean sojourn time at the FCFS node is not as simple as for the PS node. By substituting (5.9) and (5.13) in (5.15) this method gives another result than before, derived below

$$\begin{aligned} \mathrm{E}_{appr2}\{S_F\} &= p\sum_{k=1}^{\infty}(1-p)p^{k-1}((1-w)\mathrm{E}\{\sigma_F{}^{(k)}\} + wk\mathrm{E}\{\sigma_F{}^{(1)}\}) \\ &= p\sum_{k=1}^{\infty}(1-p)p^{k-1}(1-w)\mathrm{E}\{\sigma_F{}^{(k)}\} + p\sum_{k=1}^{\infty}(1-p)p^{k-1}wk\mathrm{E}\{\sigma_F{}^{(1)}\} \end{aligned}$$

$$
\begin{aligned}
&= (1-w)\frac{p}{1-p}\left(\frac{\beta_F}{1-\rho_F} + \frac{\lambda}{2}(\beta_F^{(2)} - 2\beta_F^{\,2})\frac{p}{1-\rho_F}\right) \\
&\quad + w\frac{p}{1-p}\left(\frac{\beta_F}{1-\rho_F} + \frac{\lambda}{2}(\beta_F^{(2)} - 2\beta_F^{\,2})\frac{p}{1-p}\frac{1-p\rho_F}{1-\rho_F}\right) \\
&= \frac{p}{1-p}\left(\frac{\beta_F}{1-\rho_F} + \frac{\lambda}{2}(\beta_F^{(2)} - 2\beta_F^{\,2})\frac{p}{1-\rho_F}\left(1-w+w\frac{1-p\rho_F}{1-p}\right)\right).
\end{aligned}
\tag{5.17}
$$

If the service times at the FCFS node are exponentially distributed, then $\beta_F^{(2)} - 2\beta_F^{\,2} = 0$, and thus (5.17) is exact.

Finally, the second approximation for the mean total sojourn time becomes:

$$
\begin{aligned}
\mathrm{E}_{appr2}\{S\} &\approx \frac{1}{1-p}\frac{\beta_{PS}}{1-\rho_{PS}} \\
&\quad + \frac{p}{1-p}\left(\frac{\beta_F}{1-\rho_F} + \frac{\lambda}{2}(\beta_F^{(2)} - 2\beta_F^{\,2})\frac{p}{1-\rho_F}\left(1-w+w\frac{1-p\rho_F}{1-p}\right)\right).
\end{aligned}
\tag{5.18}
$$

5.3.4 Third approximation for mean total sojourn time

The second approximation from [8], see the previous subsection, still gives rather bad results, as can be read in section 5.4. So we extend the second approximation and to improve it for some cases. Again we take an interpolation between two extreme cases, the case are the same as for the second approximation. Again, for the short-circuiting case we use just one assumption:

- $S_{PS}^{(j)}$ has the same distribution as $\sigma_{PS}^{(j)}$ and similarly $S_F^{(j)}$ has the same distribution as $\sigma_F^{(j)}$.

For the independence case, we no longer assume that customers arrive at both nodes according to a Poisson process. Instead of M/G/1 servers we assume GI/G/1 servers. In the literature there are some results for GI/G/1-PS as well as for GI/G/1-FCFS queues, these results are already mentioned in chapter 4.

For the PS node we use results from a paper by Sengupta [24]. He proposes an approximation for the sojourn time distribution for the GI/G/1-PS queue. As input we need the LST of the interarrival time distribution at the PS node and the mean service time at the PS node. The mean service time at the PS node is known and denoted by β_{PS}. For the LST of the interarrival time distribution we need the mean and SCV's of the arrival process at the PS node. Note that the arrival process at the PS node consists of external arrivals and arrivals that are fed back. So the total arrival process at the PS node is the sum of the external arrival process, for which we know the arrival rate λ, and the internal arrival process, which is equal to the departure process from the FCFS node.

Assume that we know the mean and SCV of the total arrival process at the PS node. In the appendix of Tijms [26] it is given how a hyperexponential distribution of order two (H_2) can be used for a two-moment fit. A H_2 density is a mixture of two exponentials with different means. A random variable having the H_2 density is distributed with probability p_1 as an exponential variable with mean $1/\mu_1$ and with probability $p_2 = 1 - p_1$ as an exponential variable with mean $1/\mu_2$. In general, a H_2 density has three parameters, and therefore is not uniquely determined by its first two moments.

For a two-moment fit the H_2 density with balanced means is often used. In case of balanced means, the normalization $p_1 / \mu_1 = p_2 / \mu_2$ is used. The parameters of the H_2 density having balanced means and fitting the first two moments of a positive random variable X with $c_x^2 \geq 1$ are

$$p_1 = \frac{1}{2}\left(1 + \sqrt{\frac{c_x^2 - 1}{c_x^2 + 1}}\right), \; p_2 = 1 - p_1, \tag{5.19}$$

$$\mu_1 = \frac{2p_1}{E\{X\}}, \; \mu_2 = \frac{2p_2}{E\{X\}}. \tag{5.20}$$

We use the H_2 distribution with balanced means to approximate the total arrival process at the PS node, i.e. the external arrivals and the arrivals that are fed back after visiting the FCFS node.

The LST of a H_2 distribution is given by, cf. [5]

$$\alpha(\zeta) = p_1 \frac{\mu_1}{\mu_1 + \zeta} + p_2 \frac{\mu_2}{\mu_2 + \zeta}. \tag{5.21}$$

Now let η denote the smallest positive root of the equation

$$\begin{aligned} \eta = \alpha\left(\frac{1}{\beta_{PS}}(1-\eta)\right) &= p_1 \frac{\mu_1}{\mu_1 + \frac{1}{\beta_{PS}}(1-\eta)} + p_2 \frac{\mu_2}{\mu_2 + \frac{1}{\beta_{PS}}(1-\eta)} \\ &= \frac{\frac{1}{2}\left(1 + \sqrt{\frac{c_x^2 - 1}{c_x^2 + 1}}\right)^2}{1 + \sqrt{\frac{c_x^2 - 1}{c_x^2 + 1}} + \frac{\lambda}{1-p}\frac{1-\eta}{\beta_{PS}}} + \frac{\frac{1}{2}\left(1 - \sqrt{\frac{c_x^2 - 1}{c_x^2 + 1}}\right)^2}{1 - \sqrt{\frac{c_x^2 - 1}{c_x^2 + 1}} + \frac{\lambda}{1-p}\frac{1-\eta}{\beta_{PS}}}. \end{aligned} \tag{5.22}$$

Note that one solution of (5.22) will be 1. The other two solutions follow by solving the remaining second-order equation, one of the solutions will be larger than one, and the other solution will be smaller than one. That last solution is the one we are looking for.

In section 4.3.1 Sengupta's approximation of the mean sojourn time in a PS node is given. In the notation of this chapter it follows that

$$E\{\sigma_{PS}^{(1)}\} = \frac{\beta_{PS}}{1-\eta}, \tag{5.23}$$

where η is the smallest positive root of (5.22).

Now we can write the interpolation for the PS node as follows

$$\begin{aligned} E_{appr3}\{S_{PS}\} &= \sum_{k=0}^{\infty} (1-p)p^k \left(wE\{\sigma_{PS}^{(k+1)}\} + (1-w)(k+1)E\{\sigma_{PS}^{(1)}\}\right) \\ &= \frac{1}{1-p}\left(w\frac{\beta_{PS}}{1-\rho_{PS}} + (1-w)\frac{\beta_{PS}}{1-\eta}\right). \end{aligned} \tag{5.24}$$

Note that for exponentially distributed service times at the PS node $\eta = \rho_{PS}$. Then (5.24) is equal to (5.7) and (5.18). Hence, (5.24) is an exact expression for the mean sojourn time at the PS node if the service times are exponentially distributed.

For the FCFS node we use approximations from Whitt's QNA [28]

$$E\{W_F\} = \frac{\beta_F \rho_F \left(c_{a_F}{}^2 + c_{b_F}{}^2\right) g}{2(1-\rho_F)}, \tag{5.25}$$

$$\text{where } g = \begin{cases} \exp\left[-\dfrac{2(1-\rho_F)}{3\rho_F}\dfrac{(1-c_{a_F}{}^2)^2}{c_{a_F}{}^2 + c_{b_F}{}^2}\right], & c_{a_F}{}^2 < 1 \\ 1, & c_{a_F}{}^2 \geq 1. \end{cases} \tag{5.26}$$

Whitt [28] shows how to eliminate immediate feedback. By giving an example in his QNA performance paper [29] he notes that this procedure can also be applied to almost immediate feedback, i.e. for example with one node in between. The first step of the configuring procedure is quite simple: the new service time is regarded as a geometric mixture of the n-fold convolution of the old service time distribution. Some of the parameters are changed now in the following way

$$\begin{aligned} \hat{\beta}_F &= \frac{\beta_F}{1-p}, \\ \hat{c}_{b_F}{}^2 &= p + (1-p)c_{b_F}{}^2, \\ \hat{p} &= 0. \end{aligned} \tag{5.27}$$

When we eliminate feedback in this way, we no longer count the times a customer is fed back as separate visits. So if we now calculate the mean sojourn time in the FCFS node for a customer with these new parameters (5.27), we just have to multiply it by p, because the FCFS node is not visited at all with probability $1-p$, and then we have the mean sojourn time in the FCFS node for a customer

$$\begin{aligned} p\sum_{k=1}^{\infty}(1-p)p^{k-1}k\mathrm{E}\left\{\sigma_F{}^{(1)}\right\} &= p\left(\frac{\beta_F}{1-p}\frac{\rho_F\left(c_{a_F}{}^2 + p + (1-p)c_{b_F}{}^2\right)g}{2(1-\rho_F)} + \frac{\beta_F}{1-p}\right) \\ &= \frac{p}{1-p}\left(\frac{\beta_F}{1-\rho_F} + \frac{\lambda}{2}\left(\beta_F{}^{(2)} - 2\beta_F{}^2 + \frac{\beta_F{}^2}{1-p}\left(c_{a_F}{}^2 + 1 - \frac{2}{g}\right)\right)\frac{pg}{(1-\rho_F)}\right), \end{aligned} \tag{5.28}$$

where g is as defined in (5.26).

Now we can express the interpolation for the approximation of the mean sojourn time at the FCFS node as follows

$$\begin{aligned} \mathrm{E}_{appr3}\{S_F\} &= p\sum_{k=1}^{\infty}(1-p)p^{k-1}\left((1-w)\mathrm{E}\left\{\sigma_F{}^{(k)}\right\} + wk\mathrm{E}\left\{\sigma_F{}^{(1)}\right\}\right) \\ &= (1-w)\frac{p}{1-p}\left(\frac{\beta_F}{1-\rho_F} + \frac{\lambda}{2}(\beta_F{}^{(2)} - 2\beta_F{}^2)\frac{p}{1-\rho_F}\right) + \\ &\quad w\frac{p}{1-p}\left(\frac{\beta_F}{1-\rho_F} + \frac{\lambda}{2}\left(\beta_F{}^{(2)} - 2\beta_F{}^2 + \frac{\beta_F{}^2}{1-p}\left(c_{a_F}{}^2 + 1 - \frac{2}{g}\right)\right)\frac{pg}{(1-\rho_F)}\right). \end{aligned} \tag{5.29}$$

We already mentioned that the first term is exact if the service times are exponentially distributed. If the service times are exponentially distributed and the interarrival process at the FCFS node is a Poisson process, then the second term is exact as well. This can be seen as follows. We noticed before that ${\beta_F}^{(2)} - 2{\beta_F}^2 = 0$ if the service times at the FCFS node are exponentially distributed. Further we know that if the interarrival process at the FCFS node is a Poisson process, then $\frac{{\beta_F}^2}{1-p}\left({c_{a_F}}^2 + 1 - \frac{2}{g}\right) = 0$ as well, because then ${c_{a_F}}^2 = g = 1$.

Now we need the SCV of the arrival processes at both nodes. The arrival process at the PS node is composed of external arrivals and departures from the FCFS node. The arrival process at the FCFS node consists of departures from the PS node that will be routed to the FCFS node with probability p. Note that the network in this chapter is a feedback network, hence all SCV's depends on each other.

In section 4.3.2 the SCV of interdeparture times from a PS node is given as function of the SCV of interarrival times at a PS node, that is:

$$ {c_{d_{PS}}}^2 \approx {\rho_{PS}}^2 + (1 - {\rho_{PS}}^2){c_{a_{PS}}}^2. \tag{5.30} $$

As mentioned in section 4.3.3, Whitt gives a formula for splitting. We apply Whitt's formula to our network. In our model the departure process from the PS node, with parameter ${c_{d_{PS}}}^2$, is split into two streams. Customers go to the FCFS node with probability p and leave the network with probability 1-p. The arrival process at the FCFS node now has SCV given by:

$$ {c_{a_F}}^2 \approx p{c_{d_{PS}}}^2 + 1 - p. \tag{5.31} $$

Whitt also gives a formula for the SCV of an interdeparture time in an FCFS node, see section 4.3.2:

$$ {c_{d_F}}^2 \approx {\rho_F}^2{c_{b_F}}^2 + (1 - {\rho_F}^2){c_{a_F}}^2. \tag{5.32} $$

Still we do not know the SCV of the total arrival process at the PS node. Whitt also gives formulas for superposition, see formula 4.1 in section 4.3.3. If we apply this formula to our case, we get:

$$ {c_{a_{PS}}}^2 \approx \nu\left(\frac{\lambda}{\lambda + \frac{p\lambda}{1-p}}{c_{a_0}}^2 + \frac{\left(\frac{p\lambda}{1-p}\right)}{\lambda + \frac{p\lambda}{1-p}}{c_{d_F}}^2 \right) + 1 - \nu = \nu\left((1-p){c_{a_0}}^2 + p{c_{d_F}}^2\right) + 1 - \nu, \tag{5.33} $$

where $\nu = \left[1 + 4(1-\rho_{PS})^2\left(\frac{1}{(1-p)^2 + p^2} - 1\right)\right]^{-1}$. (5.34)

Now we have equations for the SCV's of all arrival and departure processes in the network. We can solve this set of equations, such that all unknown SCV's can be calculated if the feedback probability p, the external arrival rate λ, the SCV of the external arrivals, and the loads at both nodes are known.

The resulting set of equations is

$$\begin{cases} c_{d_{PS}}{}^2 \approx \rho_{PS}{}^2 + (1-\rho_{PS}{}^2)c_{a_{PS}}{}^2 \\ c_{a_F}{}^2 \approx pc_{d_{PS}}{}^2 + 1 - p \\ c_{d_F}{}^2 \approx \rho_F{}^2 c_{b_F}{}^2 + (1-\rho_F{}^2)c_{a_F}{}^2 \\ c_{a_{PS}}{}^2 \approx v\left((1-p)c_{a_0}{}^2 + pc_{d_F}{}^2\right) + 1 - v. \end{cases} \tag{5.35}$$

Because in our model new customers arrive at the PS node according to a Poisson process, the SCV of the external arrivals, $c_{a_0}{}^2$, is equal to one. By solving this set of equations, we get an expression for the SCV of the arrival process at the FCFS node

$$c_{a_F}{}^2 \approx \frac{1-p^2 v(1-\rho_{PS}{}^2)(1-\rho_F{}^2 c_{b_F}{}^2)}{1-p^2 v(1-\rho_{PS}{}^2)(1-\rho_F{}^2)}, \tag{5.36}$$

where v is as given in (5.34).

By substitution of this expression in (5.29) for the mean sojourn time at the FCFS node we finally get an approximation of the mean sojourn time at the FCFS node in our network.

The expression for the SCV of the arrival process at the PS node is

$$c_{a_{PS}}{}^2 \approx 1 - \frac{pv\rho_F{}^2(1-c_{b_F}{}^2)}{1-p^2 v(1-\rho_{PS}{}^2)(1-\rho_F{}^2)}. \tag{5.37}$$

We can fit a H_2 distribution function for the total arrival process at the PS node by substituting $c_x{}^2 = c_{a_{PS}}{}^2$ as in (5.37) and $E\{X\} = \frac{1-p}{\lambda}$ in (5.19) and (5.20). With the LST of this distribution we can solve η, and finally with this η we can approximate the mean sojourn time in the PS node as in (5.24).

Finally, the third approximation for the mean total sojourn time in the two-node feedback network with generally distributed service times is

$$\begin{aligned} E_{appr3}\{S\} = \frac{1}{1-p}&\left(w\frac{\beta_{PS}}{1-\rho_{PS}} + (1-w)\frac{\beta_{PS}}{1-\eta}\right) \\ &+ (1-w)\frac{p}{1-p}\left(\frac{\beta_F}{1-\rho_F} + \frac{\lambda}{2}(\beta_F{}^{(2)} - 2\beta_F{}^2)\frac{p}{1-\rho_F}\right) \\ &+ w\frac{p}{1-p}\left(\frac{\beta_F}{1-\rho_F} + \frac{\lambda}{2}\left(\beta_F{}^{(2)} - 2\beta_F{}^2 + \frac{\beta_F{}^2}{1-p}\left(c_{a_F}{}^2 + 1 - \frac{2}{g}\right)\right)\frac{pg}{(1-\rho_F)}\right). \end{aligned} \tag{5.38}$$

For completeness, we refer to the definitions of the variables used in this approximation: w is given in (5.11), η is the smallest positive root of (5.22), $c_{a_F}{}^2$ is given in (5.36) and g is defined in (5.26).

Remark about general arrival process

In this chapter we assume that the external arrival process is a Poisson process. Nevertheless, we think that the third approximation is also applicable for the same network with generally distributed interarrival times. Then the set of equations (5.35) has to be solved again for the given $c_{a_0}^{\;2}$.

5.4 Numerical results

Simulation is used to validate the approximations for the mean total sojourn time in the network given in section 5.2.

We will compare the first and the second approximation from [8] to show that the second approximation gives better results than the first approximation for the case where the mean sojourn times in the PS and FCFS node are roughly equal. In fact, the second approximation is a special case of the first approximation. So the results for the cases in which one of the queues has much larger mean sojourn times at each visit than the other one should be the same for both methods. So this can be used to verify the approximations when comparing to the simulation results.

We also use simulations to validate the improvement of the approximation for the mean total sojourn time in the network, given in section 5.3.4. In the tables below we compare the simulation results to the three approximations and give the percentage error defined as in (3.30).

In the first table we give numerical results for different cases with exponentially distributed service times at both nodes. As can be seen in the table, all approximations work very well, the error is at most about 1 % and in most cases even below 0.4%. This is as expected, because the mean total sojourn time has a product-form solution in this network. We used these cases to verify that all approximations give that solution for the simple case with exponentially distributed service times, independent of the load at both nodes.

p	λ	ρ_{PS}	ρ_F	$c_{PS}^{\;2}$	$c_F^{\;2}$	Appr 1	$\Delta\%$	Appr 2	$\Delta\%$	Appr 3	$\Delta\%$
0.2	0.72	0.9	0.1	1	1	12.65	0.02	12.65	0.02	12.65	0.02
0.5	0.45	0.9	0.1	1	1	20.25	1.03	20.25	1.03	20.25	1.03
0.8	0.18	0.9	0.1	1	1	50.62	-0.30	50.62	-0.30	50.62	-0.30
0.2	0.08	0.1	0.9	1	1	113.89	-0.08	113.89	-0.08	113.89	-0.08
0.5	0.05	0.1	0.9	1	1	182.22	0.35	182.22	0.35	182.22	0.35
0.8	0.04	0.1	0.9	1	1	227.78	-0.18	227.78	-0.18	227.78	-0.18
0.2	0.64	0.8	0.8	1	1	12.50	-0.38	12.50	-0.38	12.50	-0.38
0.5	0.4	0.8	0.8	1	1	20.00	0.35	20.00	0.35	20.00	0.35
0.8	0.16	0.8	0.8	1	1	50.00	0.26	50.00	0.26	50.00	0.26

Table 5.1 Approximations for total mean sojourn time in feedback network with a PS node and a FCFS node, with exponential service times at both nodes

Now we know that our simulation model is valid and our approximations work well for the simple case with exponentially distributed service times. A next step is applying the approximations to networks with generally distributed service times. We first use hyperexponentially distributed service times with balanced means, but the method is also applicable to other service time distributions.

In the second table we illustrate the remark of Boxma et al. that the second assumption for the first approximation works well when one of the two queues has much larger mean sojourn times at each visit than the other. If the load at one of the nodes is much higher than at the other node in the network, then the mean sojourn time at that first node will be larger than the mean sojourn time at the other node in the network too. This is true for the cases shown in table 5.2.

For the cases shown in table 5.2 all three approximations work well. The relative error is still small, smaller than 2% in all cases, and even smaller than 1% in most cases. Note that the relative error is sometimes negative and sometimes positive. This means that it is not the case that the approximation always underestimates or overestimates the mean total sojourn time. So we can conclude that all three approximations work well for the cases given in tables 5.1 and 5.2.

λ	ρ_{PS}	ρ_F	$c_{b_{PS}}^2$	$c_{b_F}^2$	Sim	Appr 1	$\Delta\%$	Appr 2	$\Delta\%$	Appr 3	$\Delta\%$
0.36	0.9	0.1	1.67	1.67	25.16	25.31	0.61	25.32	0.64	25.32	0.65
0.09	0.9	0.1	1.67	4.56	100.82	101.28	0.46	101.43	0.61	101.47	0.65
0.36	0.9	0.1	4.56	4.56	25.60	25.32	-1.10	25.36	-0.95	25.37	-0.91
0.09	0.9	0.1	4.56	1.67	103.18	101.24	-1.88	101.27	-1.85	101.28	-1.85
0.01	0.1	0.9	1.67	1.67	958.54	965.11	0.69	965.31	0.71	965.74	0.75
0.04	0.1	0.9	1.67	4.56	302.01	299.78	-0.74	299.97	-0.67	300.41	-0.53
0.04	0.1	0.9	4.56	4.56	299.44	299.78	0.11	299.97	0.18	300.41	0.32
0.01	0.1	0.9	4.56	1.67	951.23	965.11	1.46	965.31	1.48	965.74	1.53

Table 5.2 Approximations for total mean sojourn time in feedback network with a PS node and a FCFS node, with general service times at both nodes

As mentioned before, the cases in given in tables 5.1 and 5.2 are not the real problem. The second approximation is an improvement of the first approximation for cases where the mean sojourn times at both queues are roughly equal. In table 5.3 we show some cases where the load at the two nodes is equal. As a result the mean sojourn times at both nodes will be roughly equal. Again we give the values for all three approximations and compare these with the simulation results. Here ρ denotes $\rho_{PS} = \rho_F$.

The second approximation is always closer to the simulation result than the first one. However, as can be seen, the second approximation still gives bad results in most cases where the feedback probability is rather high. From these, and much more, numerical results, we concluded that the development of the third approximation, see section 5.3.4, was needed for cases with a high feedback probability, equal load at both nodes, and non-exponentially distributed service times at at least the FCFS node. For most cases the third approximation is very close to the simulation value, in one case in the table the third approximation is worse than the second.

Nr.	p	λ	ρ	$c_{b_{PS}}^2$	$c_{b_F}^2$	Sim	Appr 1	$\Delta\%$	Appr 2	$\Delta\%$	Appr 3	$\Delta\%$
1	0.5	0.20	0.80	1.67	1.67	44.09	42.67	-3.24	42.83	-2.87	44.26	0.37
2	0.5	0.80	0.80	1.67	1.67	11.05	10.67	-3.50	10.71	-3.14	11.06	0.09
3	0.8	0.08	0.80	1.00	4.56	154.08	114.22	-25.87	118.16	-23.32	153.73	-0.23
4	0.8	0.10	0.50	1.00	4.56	24.55	21.78	-11.28	23.00	-6.31	23.60	-3.87
5	0.8	0.10	0.50	4.56	4.56	23.48	21.78	-7.25	23.00	-2.05	23.60	0.50
6	0.8	0.10	0.50	1.67	4.56	24.21	21.78	-10.04	23.00	-5.01	23.60	-2.53
7	0.8	0.08	0.80	1.67	1.67	109.55	102.67	-6.29	103.57	-5.46	110.34	0.72
8	0.8	0.32	0.80	1.67	1.67	27.44	25.67	-6.47	25.89	-5.65	27.58	0.52
9	0.8	0.08	0.80	1.67	4.56	145.18	114.22	-21.33	118.16	-18.61	153.73	5.89
10	0.8	0.08	0.80	4.56	4.56	133.10	114.22	-14.18	118.16	-11.22	153.73	15.50
11	0.8	0.32	0.80	4.56	1.67	26.81	25.67	-4.26	25.89	-3.42	27.58	2.90
12	0.8	0.08	0.80	4.56	1.67	107.95	102.67	-4.90	103.57	-4.06	110.34	2.21

Table 5.3 Approximations for total mean sojourn time in feedback network with a PS node and a FCFS node, with general service times at both nodes and equal load at both nodes.

We want to know if the third approximation gives better or at least good results for the mean sojourn times at both nodes separately as well. So below we give the results for the mean sojourn time at the PS node and the FCFS node separately. Note that the case numbers in table 5.4 and table 5.5 are the same as in table 5.3. So the characteristics of the simulated networks in table 5.4 and table 5.5 can be found in table 5.3.

Nr.	$c_{b_{PS}}^{2}$	Appr $c_{a_{PS}}^{2}$	Sim	Appr 1	Δ%	Appr 2	Δ%	Appr 3	Δ%
1	1.67	1.19	21.12	20.00	-5.31	20.00	-5.31	21.05	-0.32
2	1.67	1.19	5.27	5.00	-5.17	5.00	-5.17	5.26	-0.18
3	1.00	2.83	76.22	50.00	-34.40	50.00	-34.40	73.45	-3.64
4	1.00	1.64	11.97	10.00	-16.45	10.00	-16.45	10.98	-8.27
5	4.56	1.64	11.10	10.00	-9.94	10.00	-9.94	10.98	-1.13
6	1.67	1.64	11.71	10.00	-14.57	10.00	-14.57	10.98	-6.21
7	1.67	1.34	54.25	50.00	-7.83	50.00	-7.83	53.95	-0.55
8	1.67	1.34	13.59	12.50	-7.99	12.50	-7.99	13.49	-0.71
9	1.67	2.83	70.46	50.00	-29.04	50.00	-29.04	73.45	4.24
10	4.56	2.83	61.68	50.00	-18.94	50.00	-18.94	73.45	19.07
11	4.56	1.34	13.12	12.50	-4.75	12.50	-4.75	13.49	2.78
12	4.56	1.34	52.53	50.00	-4.81	50.00	-4.81	53.95	2.71

Table 5.4 Approximations for the mean sojourn time in the PS node, the numbers correspond to the numbers in table 5.3

Table 5.4 shows that the third approximation gives better results for the mean sojourn time at the PS node than the first and second approximations do. Note that all three approximations of the mean sojourn time at the PS node do not distinguish between the second moment of the service time at the PS node. Compare, e.g., case number 3, 9, and 10 with different SCV's for the service times. For the first and second approximation we assume that the PS node is a M/G/1-PS node. For the third approximation we used Sengupta's approximation, see section 4.4.1. In that section we give some remarks about Sengupta's approximation. It is probably due to the combination of a large SCV for the service time and a large approximate SCV for the interarrival process that the third approximation is still bad for the tenth case.

Nr.	$c_{b_F}^{2}$	Appr $c_{a_F}^{2}$	Sim	Appr 1	Δ%	Appr 2	Δ%	Appr 3	Δ%
1	1.67	1.03	22.97	22.67	-1.33	22.83	-0.63	22.75	-0.97
2	1.67	1.03	5.78	5.67	-1.98	5.71	-1.29	5.69	-1.63
3	4.56	1.53	77.86	64.222	-17.52	68.159	-12.46	67.88	-12.82
4	4.56	1.38	12.58	11.778	-6.36	12.997	3.34	12.11	-3.74
5	4.56	1.38	12.38	11.78	-4.83	13.00	5.03	12.11	-2.16
6	4.56	1.38	12.50	11.78	-5.80	13.00	3.95	12.11	-3.17
7	1.67	1.10	55.30	52.67	-4.77	53.57	-3.14	53.50	-3.25
8	1.67	1.10	13.86	13.17	-4.98	13.39	-3.35	13.38	-3.47
9	4.56	1.53	74.72	64.22	-14.05	68.16	-8.78	67.88	-9.16
10	4.56	1.53	71.42	64.22	-10.07	68.16	-4.56	67.88	-4.95
11	1.67	1.10	13.69	13.17	-3.79	13.39	-2.14	13.38	-2.26
12	1.67	1.10	55.43	52.67	-4.98	53.57	-3.35	53.50	-3.47

Table 5.5 Approximations for the mean sojourn time in the FCFS node, the numbers correspond to the numbers in table 5.3

Table 5.5 shows that the third approximation for the mean sojourn time at the FCFS node does not give a significant improvement compared to the second approximation. For the second approximation we assumed that the FCFS node is an M/G/1-FCFS node, for the third approximation we assumed a GI/G/1-FCFS node. The SCV for the arrival process is also approximated. The approximations for these squared coefficients are given in the table. These approximated coefficients are not compared to the simulated values, this is an interesting topic for further research. We do not know much about the influence of the approximations for the SCV's yet. Although in chapter 4 the accuracy of the SCV approximations did not seem to be important for getting accurate approximations of the mean sojourn times. But intuitively, it is also possible that inaccuracies in the approximations for the SCV's in the network lead to large errors in the mean sojourn time approximations.

Gamma distributed service times

In this chapter, we developed approximations for a network with generally distributed service times. By applying the developed approximations to a network with hyperexponentially distributed service times, we showed that the approximations are accurate for that case. We applied the approximations to a network with gamma distributed service times as well. Instead of the H_2 fit we fit a gamma distribution on the interarrival times at the PS node.

The parameters of the gamma density fitting the first two moments of a positive random variable X with $c_x^2 \geq 1$ are

$$\varphi = \frac{1}{c_x^2}, \tag{5.39}$$

$$\mu = \frac{\varphi}{E\{X\}}. \tag{5.40}$$

The LST of a gamma distribution is given by, cf. [5]

$$\alpha(\zeta) = \frac{\mu^{\varphi}}{(\mu+\zeta)^{\varphi}}. \tag{5.41}$$

Now we can approximate the mean total sojourn time in the network by using (5.38), let η denote the smallest positive root of the equation

$$\eta = \alpha\left(\frac{1}{\beta_{PS}}(1-\eta)\right) = \left(\frac{\dfrac{1}{c_{a_{PS}}^{2}\dfrac{1-p}{\lambda}}}{\dfrac{1}{c_{a_{PS}}^{2}\dfrac{1-p}{\lambda}} + \dfrac{1-\eta}{\beta_{PS}}}\right)^{\frac{1}{c_{a_{PS}}^{2}}}. \tag{5.42}$$

In table 5.6 below we give results for gamma distributed service times with the SCV's of the service time distributions at both nodes smaller than or equal to one. The approximation works for SCV's of service times larger than one as well, we already showed this for the hyperexponential distribution above.

p	λ	ρ_{PS}	$c_{b_{PS}}^2$	ρ_F	$c_{b_F}^2$	Sim	Appr 1	$\Delta\%$	Appr 2	$\Delta\%$	Appr 3	$\Delta\%$
0.5	0.80	0.8	0.50	0.8	1.00	10.05	10.00	-0.46	10.00	-0.46	10.00	-0.46
0.8	0.20	0.8	0.50	0.8	1.00	40.54	40.00	-1.33	40.00	-1.33	40.00	-1.33
0.5	0.40	0.8	1.00	0.8	0.50	18.26	19.00	4.07	18.93	3.67	17.42	-4.61
0.8	0.16	0.8	1.00	0.8	0.50	45.14	49.00	8.54	48.63	7.72	42.70	-5.42
0.5	0.40	0.8	1.00	0.8	0.25	17.01	18.50	8.75	18.39	8.08	16.47	-3.19
0.8	0.16	0.8	1.00	0.8	0.25	42.68	48.50	13.63	47.93	12.30	40.16	-5.92
0.5	0.45	0.9	1.00	0.9	0.25	33.50	36.63	9.31	36.50	8.93	32.54	-2.88
0.8	0.18	0.9	1.00	0.9	0.25	83.69	96.63	15.45	95.99	14.70	80.17	-4.21
0.5	0.25	0.5	1.00	0.8	0.25	16.84	17.60	4.54	17.54	4.16	16.82	-0.09
0.8	0.10	0.5	1.00	0.8	0.25	44.47	47.60	7.03	47.25	6.24	43.97	-1.13
0.5	0.10	0.2	1.00	0.5	0.25	11.30	11.56	2.34	11.50	1.82	11.09	-1.84
0.8	0.04	0.2	1.00	0.5	0.25	29.45	30.31	2.94	29.97	1.77	28.30	-3.89

Table 5.6 Approximations for total mean sojourn time in feedback network with a PS node and a FCFS node, with gamma distributed service times at both nodes.

In table 5.6 we show that for several parameter combinations our third approximation for the mean total sojourn time is rather accurate. We show some interesting cases, such as equal loads at both nodes. As one can see, the third approximation is for most cases closer to the simulated value than the second approximation. If the second approximation is better than the third approximation, then the third approximation is an accurate approximation as well. It is remarkable that the third approximation underestimates the simulated values in all cases considered in table 5.6. We think that the used approximation for the mean waiting time in the FCFS node is an underestimation. This is probably due to the elimination of feedback, we applied formulas for eliminating direct feedback to a network with indirect feedback. Note that Whitt [29] remarks that his approximation works well if the percentage error is, e.g., about 10%. So we conclude that our approximation works well too, resulting in a percentage error of maximum about 6% for the mean total sojourn time in the considered cases, with one exception of 15%.

Chapter 6

Summary and further research

6.1 Summary of the results

In this thesis we obtain approximations for the mean as well as the variance of the total sojourn time in a feedback network with a single PS node and several multi-server FCFS nodes. In this network the interarrival times as well as the service times are exponentially distributed. We extend the network in [14] to a network with multi-server instead of single-server FCFS nodes.

We also obtain an approximation for the mean total sojourn time in a feedforward network with a PS node and a single-server FCFS node. In this network the interarrival times as well as the service times at both nodes are generally distributed. Two-moment approximations for mean sojourn times in GI/G/1 nodes are used to approximate the mean sojourn times in the networks considered. An important result in this chapter is an approximate expression for the squared coefficient of variation of the departure process from a PS node.

Finally we improve an approximation for the mean total sojourn time in a feedback network with a PS node and a single-server FCFS node. In this network the external arrival process is a Poisson process and the service times are generally distributed. Besides validating approximations developed by Boxma et al. [8], these approximations for the mean total sojourn time in the network are improved.

We validated all approximations by comparing approximations with simulation values. Therefore we run a lot of simulations. We conclude that the resulting approximations obtained in this thesis are all explicit, accurate, and quite fast-to-evaluate approximations.

6.2 Topics for further research

The results for approximations for mean total sojourn times in networks with general service times can be extend for networks with multiple multi-sever FCFS nodes, as discussed in chapter 3. To this end, the results from chapter 3 and 5 have to be combined, which addresses an interesting topic for further research.

The networks under consideration can also be extended by taking into account multiple PS nodes.

In this thesis we mainly concentrate on the mean of the total sojourn time in a network. A topic for further research is to extend the current approximations for the mean and variances of the total sojourn time to the complete sojourn time distribution.

Another interesting extension would be to obtain approximations for the sojourn time in networks with deterministic instead of Bernoulli feedback, which significantly enhances the application possibilities of the model.

Another potential model extension is to include multiple customer types that may each be governed by different feedback schemes.

References

[1] Albin, S.L. *Approximating Queues with Superposition Arrival Processes*. Ph.D dissertation, Department of Industrial Engineering and Operations Research, Columbia University, 1981.

[2] Albin, S.L. Approximating a Point Process by a Renewal Process,II: Superposition Arrival Processes to Queues, Department of Industrial Engineering, Rutgers University, 1982.

[3] Baskett, F., K.M. Chandy, R.R. Muntz, F.G. Palacios. Open, closed, and mixed networks of queues with different classes of customers, *Journal of the Association for Computing Machinery* **22** (1975), 248-260.

[4] Van den Berg, J.L. *Sojourn Times in Feedback and Processor Sharing Queues*. Ph.D Thesis, University of Utrecht, 1990.

[5] Blanc, J.P.C. *Queueing models, analytical and numerical methods*. Tilburg University, 2004.

[6] Boucherie, R.J. *Product-form in Queueing Networks*. Ph.D. Thesis, Vrije Universiteit Amsterdam, 1992.

[7] Boxma, O.J., H. Daduna. Sojourn times in queueing networks. In: *Stochastic Analysis of Computer and Communication Systems*, ed. H. Takagi (North-Holland Publ. Cy., Amsterdam, 1990), 401-450.

[8] Boxma, O.J., B.M.M. Gijsen, R.D. van der Mei, J.A.C. Resing, Sojourn time approximations in two-node queueing networks, *Proc. HETNET-2004* (Bradford, UK, 2004).

[9] Coffman, E.G., R.R. Muntz and H. Trotter, Waiting time distributions for processor-sharing systems, *J. Assoc. Comp. Mach.* **17** (1970), 123-130.

[10] Cohen, J.W., *The single server queue*. North-Holland Publ. Cy., Amsterdam, second edition 1982.

[11] Disney, R.L., and D. Koenig, Queueing networks: a survey of their random processes, *SIAM Rev.* **27** (1985), 335-403.

[12] Doshi, B.T.and J.S. Kaufman, Sojourn time in an M/G/1 queue with Bernoulli feedback. In *Queueing Theory and its Applications-Liber Amicorum for J.W. Cohen*, eds. O.J. Boxma and R. Syski (North-Holland Publ. Cy., Amsterdam, 1988), 207-233.

[13] Engelberts, P., *Sojourn times in a queueing network with feedback.* Master's thesis, Erasmus University Rotterdam, 2003.

[14] Gijsen, B.M.M., R.D. van der Mei, P. Engelberts, J.L van den Berg. Response times in a queueing network with feedback, *submitted* (2003).

[15] Jackson, J.R. Networks of waiting lines, *Oper. Res.* **5** (1957), 518-121.

[16] Jagerman, D.L., B. Sengupta, The GI/M/1 processor-sharing queue and its heavy traffic analysis, *Commun. Statist.-stochastic models,* **7(3)** (1991), 379-395.

[17] Kelly, F.P., Networks of queues with customers of different types, *J. Appl. Prob.* **12** (1975), 542-554.

[18] Kraemer, W. and M. Langenbach-Belz, Approximate Formulae for the Delay in the Queueing System GI/G/1, *Congressbook*, Eight Int. Teletraffic Cong., Melbourne (1976), 235-1/8.

[19] Law, A.M., W.D. Kelton. *Simulation Modelling and Analysis*. McGraw-Hill, Singapore, third edition, 2000.

[20] Van der Mei, R.D., B.M.M. Gijsen, N. in 't Veld, J.L. van den Berg. Response times in a two-node queueing network with feedback, *Performance Evaluation* **49** (2002), 99-110.

[21] Morrison, J.A., Response-time distribution for a processor-sharing system, *SIAM J. App Math.* **45** (1985),152-167.

[22] Ott, T.J., The sojourn time distribution in the M/G/1 queue with processor sharing, *J. Appl. Prob.* **21** (1984), 360-378.

[23] Ramaswami, V., The sojourn time in the GI/M/1 queue with processor sharing, *J. Appl Prob.* **21** (1984), 437-442

[24] Sengupta, B., An approximation for the sojourn-time distribution for the GI/G/1 processor-sharing queue, *Commun. Statist.-stochastic models,* **8(1)** (1992), 35-57.

[25] Takács, L., A single-server queue with feedback, *The Bell System Technical Journal*, **42** (1963), 505-519.

[26] Tijms, H.C., *Stochastic Models: an algorithmic approach*. Wiley, Chichester, 1993.

[27] In 't Veld, N., *Sojourn times in a two-node queueing network with feedback*. Master's thesis, Vrije Universiteit Amsterdam, 2002.

[28] Whitt, W., The Queueing Network Analyzer, *The Bell System Technical Journal*, **62** (1983), 2779-2815.

[29] Whitt, W., Performance of the Queueing Network Analyzer, *The Bell System Technical Journal*, **62** (1983), 2817-2843.

[30] Wolff, R.W., *Stochastic modeling and the theory of queues.* Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1989.

Appendices

Appendix A

Simulations in Extend

We used the computer package Extend (see www.imaginethatinc.com) to simulate our queueing networks. Extend is a general-purpose simulation program. We used it for the validation of analytic expressions for the sojourn time in a queueing network.

With Extend it is possible to create a block diagram of a process where each block describes one part of the process. In Extend the process is laid out in a two-dimensional drawing environment and Extend has the possibility to show animation, so that customers can be seen while moving through the network. It is possible to create models quickly because Extend comes with the blocks that are needed for most simulations. The models can be built without even having to type an equation. But it is also possible to create custom blocks for specialized applications.

For the verification of the simulation models we used different techniques. First of all we always started to build a simple model, e.g. an isolated PS node, and run that model, before building a final model. The simplified model's output can be compared with its true characteristics to verify this model.

For a final model it is also possible to verify the model by comparing its output to its true characteristics. For example, if the feedback probability in a feedback network is set to zero, then the model is no longer a feedback model, but becomes a feedforward network. Thus the results can be compared with known results for a feedforward network. Another example is setting the load close to one. Often, heavy-traffic expressions are known.

For the calculation of the mean and variance of the sojourn times in our networks in Extend, we calculate the total sojourn time of every customer by subtracting its start time from its end time. We use the standard functions in Extend to calculate the mean and variance of all sojourn times. For higher-order moments long runs are needed, cf. Law and Kelton [19].

For each parameter case we ran ten simulation runs and calculated the means over the ten runs for the output parameters.

For our research we need the network in steady state, however, the simulation network always starts empty. Therefore we did not measure the performance measures during a so-called warm-up period for every run. The warm-up period was chosen long enough to guarantee that the model was in steady state.

For the calculation of confidence intervals we used formula (4.12) on page 255 of Law and Kelton [19]

On the next page an example of a simulation model in Extend is given. It is a feedback network with a PS node and a single-server FCFS node. The feedback mechanism is Bernoulli feedback. The green balls represents customers in the network. We measure the total sojourn time in the network by subtracting the start time from the end time of a customer in the network. Next we calculate the mean and variance of the total sojourn times in the network and send these performance measures for every run to a file.

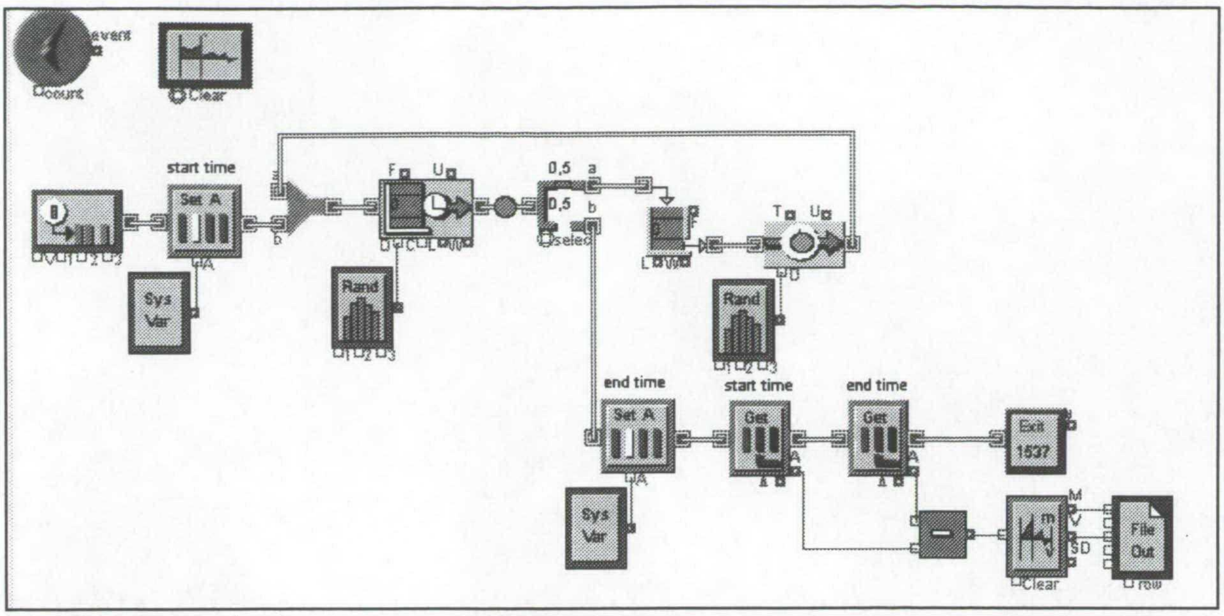


Figure A1 Extend model

Appendix B

Supplement to subsection 3.3.2

Here we give the complete derivation of the approximation for the variance of the total sojourn time, the main part of it can also be found in section 4.3.2.

Like in [14] we rewrite the variance of the sojourn time in the following way

$$Var\{S\} = Var\left\{\sum_{i=1}^{N+1} S_{PS}^{i} + \sum_{k=1}^{M}\sum_{i=1}^{N_k} S_{F_k}^{i}\right\} \tag{B1}$$

$$= E\left\{Var\left\{\sum_{i=1}^{N+1} S_{PS}^{i} + \sum_{k=1}^{M}\sum_{j=1}^{N_k} S_{F_k}^{j} \mid N_1,\dots,N_M\right\}\right\}$$
$$+ Var\left\{E\left\{\sum_{i=1}^{N+1} S_{PS}^{i} + \sum_{k=1}^{M}\sum_{j=1}^{N_k} S_{F_k}^{j} \mid N_1,\dots,N_M\right\}\right\} \tag{B2}$$

$$= \sum_{n_1=0}^{\infty}\dots\sum_{n_M=0}^{\infty} Var\left\{\sum_{i=1}^{n+1} S_{PS}^{i} + \sum_{k=1}^{M}\sum_{j=1}^{n_k} S_{F_k}^{j}\right\} f(n_1,\dots,n_M)$$
$$+ Var\left\{\sum_{i=1}^{N+1} E\{S_{PS}^{i}\} + \sum_{k=1}^{M}\sum_{j=1}^{N_k} E\{S_{F_k}^{j}\}\right\} \tag{B3}$$

$$= \sum_{n=0}^{\infty} Var\left\{\sum_{i=1}^{n+1} S_{PS}^{i}\right\}(1-p)p^{n} + \sum_{k=1}^{M}\sum_{n_k=0}^{\infty} Var\left\{\sum_{j=1}^{n_k} S_{F_k}{}^{j}\right\}(1-q_k)q_k{}^{n_k} +$$
$$2\sum_{k=1}^{M}\sum_{n_1=0}^{\infty}\dots\sum_{n_M=0}^{\infty} Cov\left\{\sum_{i=1}^{n+1} S_{PS}^{i}, \sum_{j=1}^{n_k} S_{F_k}^{j}\right\} f(n_1,\dots,n_M) +$$
$$\sum_{k\neq m}\sum_{n_1=0}^{\infty}\dots\sum_{n_M=0}^{\infty} Cov\left\{\sum_{i=1}^{n_k} S_{F_k}^{i}, \sum_{i=1}^{n_m} S_{F_m}^{i}\right\} f(n_1,\dots,n_M) +$$
$$Var\left\{\sum_{i=1}^{N+1} E\{S_{PS}^{i}\} + \sum_{k=1}^{M}\sum_{j=1}^{N_k} E\{S_{F_k}^{j}\}\right\} \tag{B4}$$

$$= \sum_{n=0}^{\infty}(n+1)Var\{S_{PS}^{1}\}(1-p)p^{n} + \sum_{n=0}^{\infty}\sum_{i\neq j} Cov\{S_{PS}^{i}, S_{PS}^{j}\}(1-p)p^{n} +$$
$$\sum_{k=1}^{M}\sum_{n_k=0}^{\infty} n_k Var\{S_{F_k}{}^{1}\}(1-q_k)q_k{}^{n_k} + \sum_{k=1}^{M}\sum_{n_k=0}^{\infty}\sum_{i\neq j} Cov\{S_{F_k}^{i}, S_{F_k}^{j}\}(1-q_k)q_k{}^{n_k} +$$
$$2\sum_{k=1}^{M}\sum_{n_1=0}^{\infty}\dots\sum_{n_M=0}^{\infty} Cov\left\{\sum_{i=1}^{n+1} S_{PS}^{i}, \sum_{j=1}^{n_k} S_{F_k}^{j}\right\} f(n_1,\dots,n_M) +$$
$$\sum_{k\neq m}\sum_{n_1=0}^{\infty}\dots\sum_{n_M=0}^{\infty} Cov\left\{\sum_{j=1}^{n_k} S_{F_k}^{i}, \sum_{j=1}^{n_m} S_{F_m}^{i}\right\} f(n_1,\dots,n_M) +$$
$$Var\left\{\sum_{k=1}^{M} N_k\left(\{S_{PS}^{1}\} + E\{S_{F_k}^{1}\}\right)\right\}. \tag{B5}$$

The first step follows from the definition of the total sojourn time. The second step follows directly from the classical formula for rewriting a variance: $Var\{X\} = E\{Var\{X \mid Y\}\} + Var\{E\{X \mid Y\}\}$. The next equation is then obtained by conditioning with respect to the event $\{N_1 = n_1,\ldots, N_M = n_M\}$. Then we use the classical formula: $Var\{\sum(X)\} = \sum(Var\{X\})$ .The final equation follows from the fact that the successive sojourn times $S_{PS}^{i}, i = 1,\ldots N+1$ are identically distributed, similarly the successive sojourn times $S_{F_k}^{j}, j = 1,\ldots N_k$ are identically distributed for each $k = 1,\ldots,M$.

Since the number of visits to the k^{th} FCFS node, N_k, is geometrically distributed with parameter q_k, we know that

$$\sum_{k=1}^{M} Var\{N_k\} = \sum_{k=1}^{M} \frac{q_k}{(1-q_k)^2}. \tag{B6}$$

In [13] the variation of the number of returns to the PS node is written in the following way

$$\begin{aligned} Var(N) &= Var\left(\sum_{k=1}^{M} N_k\right) = Var\left(\sum_{k=1}^{M-1} N_k\right) + Var(N_M) + 2Cov\left(\sum_{k=1}^{M-1} N_k, N_M\right) = \\ &= Var\left(\sum_{k=1}^{M-1} N_k\right) + Var(N_M) + 2\sum_{k=1}^{M-1} Cov(N_k, N_M) = \\ &= \sum_{k=1}^{M} Var(N_k) + \sum_{k \neq l} Cov(N_k, N_l). \end{aligned} \tag{B7}$$

Next it is obtained that

$$\sum_{k \neq m} Cov\{N_k, N_m\} = Var\{N\} - \sum_{k=1}^{M} Var\{N_k\} = \frac{p}{(1-p)^2} - \sum_{k=1}^{M} \frac{q_k}{(1-q_k)^2} = \sum_{k \neq m} \frac{p_k p_m}{(1-p)^2}. \tag{B8}$$

By substitution, we obtain

$$\begin{aligned} & Var\left\{\sum_{i=1}^{N+1} E\{S_{PS}^{i}\} + \sum_{k=1}^{M} \sum_{i=1}^{N_k} E\{S_{F_k}^{i}\}\right\} = Var\left\{\sum_{k=1}^{M} N_k \left(E\{S_{PS}^{1}\} + E\{S_{F_k}^{1}\}\right)\right\} \\ &= \sum_{k=1}^{M} Var\{N_k\}\left(E\{S_{PS}^{1}\} + E\{S_{F_k}^{1}\}\right)^2 + \\ &\quad \sum_{k \neq m} Cov\{N_k, N_m\}\left(E\{S_{PS}^{1}\} + E\{S_{F_k}^{1}\}\right)\left(E\{S_{PS}^{1}\} + E\{S_{F_m}^{1}\}\right) \\ &= \sum_{k=1}^{M} \frac{q_k}{(1-q_k)^2}\left(E\{S_{PS}^{1}\} + E\{S_{F_k}^{1}\}\right)^2 + \\ &\quad \sum_{k \neq m} \frac{p_k p_m}{(1-p)^2}\left(E\{S_{PS}^{1}\} + E\{S_{F_k}^{1}\}\right)\left(E\{S_{PS}^{1}\} + E\{S_{F_m}^{1}\}\right). \end{aligned} \tag{B9}$$

Here we know, see section 3.3.1:

$$E\{S_{PS}{}^{1}\} = \frac{\beta_{PS}}{1-\rho_{PS}}, \tag{B10}$$

$$E\{S_{F_k}{}^1\} = \beta_{F_k} + \frac{p_{w_k}}{c_k}\frac{\beta_{F_k}}{1-\rho_{F_k}}. \tag{B11}$$

Now we still need formulas for $Var\{S_{PS}^1\}, Cov\{S_{PS}^i, S_{PS}^j\}, Var\{S_{F_k}^1\}, Cov\{S_{F_k}^i, S_{F_k}^j\}, Cov\{S_{PS}^i, S_{F_k}^j\}$ and $Cov\{S_{F_k}^i, S_{Fm}^i\}$, for any $i, j = 1,2,..., k \neq m$.

Under the first approximation assumption, see section 4.3.2, we can use Ott's result [22] for the variance of the sojourn time in an M/M/1-PS system.

$$Var\{S_{PS}^1\} = \frac{2+\rho_{PS}}{2-\rho_{PS}}\left(\frac{\beta_{PS}}{1-\rho_{PS}}\right)^2. \tag{B12}$$

In [27] it is given that

$$Cov\{S_{PS}^i, S_{PS}^{i+l}\} = \frac{\rho_{PS}\beta_{PS}{}^2}{(1-\rho_{PS})^2(2-\rho_{PS}-p+\rho_{PS}p)^{l+1}}. \tag{B13}$$

Then by substitution it follows that

$$\begin{aligned}
&\sum_{n=0}^{\infty}\sum_{i\neq j} Cov\{S_{PS}^i, S_{PS}^j\}(1-p)p^n \\
&= \sum_{n=0}^{\infty}\sum_{l=1}^{n} 2(n+1-l)\frac{\rho_{PS}\beta_{PS}{}^2}{(1-\rho_{PS})^2(2-\rho_{PS}-p+p\rho_{PS})^{l+1}}(1-p)p^n \\
&= 2\frac{\rho_{PS}\beta_{PS}{}^2(1-p)}{(1-\rho_{PS})^2}\sum_{l=1}^{\infty}\left(\frac{1}{2-\rho_{PS}-p+p\rho_{PS}}\right)^{l+1}\sum_{n=l}^{\infty}(n+1-l)p^n \\
&= 2\frac{\rho_{PS}\beta_{PS}{}^2(1-p)}{(1-\rho_{PS})^2}\sum_{l=1}^{\infty}\left(\frac{p}{2-\rho_{PS}-p+p\rho_{PS}}\right)^{l+1}\frac{1}{(1-p)^2} \\
&= \frac{2\rho_{PS}\beta_{PS}{}^2}{(1-\rho_{PS})^2}\frac{p(1-p)}{(1-p)^2}\frac{\left(\frac{1}{2-\rho_{PS}-p+p\rho_{PS}}\right)^2}{1-\frac{p}{2-\rho_{PS}-p+p\rho_{PS}}} \\
&= \frac{2\rho_{PS}\beta_{PS}{}^2}{(1-\rho_{PS})^2}\frac{p}{(1-p)}\left(\frac{1}{2-\rho_{PS}-p+p\rho_{PS}}\right)^2\frac{2-\rho_{PS}-p+p\rho_{PS}}{2-\rho_{PS}-2p+p\rho_{PS}} \\
&= \frac{2p\rho_{PS}\beta_{PS}{}^2}{(1-\rho_{PS})^2(1-p)^2(2-\rho_{PS}-p+p\rho_{PS})(2-\rho_{PS})}.
\end{aligned} \tag{B14}$$

The variance of the sojourn time in the k^{th} FCFS node can be derived from the theory about M/M/c queues in [5]. Here the distribution of the sojourn time of a customer in a multi-server FCS-node is obtained as the convolution of the waiting time and the service time distribution. This leads to the following expression for the sojourn time distribution in a multi-server FCFS-node

$$P\{S \le y\} = 1 - e^{-\mu y} + \frac{p_w}{1 - c(1-\rho)}\left[e^{-\mu y} - e^{-c\mu(1-\rho)y}\right], y \ge 0. \quad \text{(B15)}$$

This is a Cox C_2 distribution consisting of an exponentially distributed service phase and an exponentially distributed waiting phase. With probability $1 - p_w$ the sojourn time is exponentially distributed with rate μ, and with probability p_w the sojourn time is exponentially distributed with rate $c\mu(1-\rho)$. The moments of a Cox C_ψ distribution are given in [5]. We can obtain the variance of the sojourn time in a multi-server FCFS-node and apply this to our case

$$Var\{S_{F_k}^1\} = \beta_{Fk}^2 + \frac{p_w(2-p_w)\beta_{Fk}^2}{c_k^2(1-\rho_{Fk})^2}. \quad \text{(B16)}$$

In [27] it is also given that

$$Cov\{S_{F_k}^i, S_{F_k}^{i+l}\} = \frac{\rho_{F_k}(\rho_{F_k}(1-p)+p)^{l-1}\beta_{F_k}^2}{(1-\rho_{F_k})^2}. \quad \text{(B17)}$$

This does not directly hold for a multi-server FCFS node. So we need to develop an approximation for the $Cov\{S_{F_k}^i, S_{F_k}^{i+l}\}$ in case of a model with multi-server FCFS nodes instead of single-server FCFS nodes. The correlation between $S_{F_k}^i$ and $S_{F_k}^{i+l}$ is defined as

$$Corr\{S_{F_k}^i, S_{F_k}^{i+l}\} = \frac{Cov\{S_{F_k}^i, S_{F_k}^{i+l}\}}{\sqrt{Var\{S_{F_k}^i\}}\sqrt{Var\{S_{F_k}^{i+l}\}}}. \quad \text{(B18)}$$

As an approximation for this correlation we can take an interpolation between light and heavy traffic. The probability that the system is in a heavy traffic situation is equal to the probability that a customer has to wait; this is p_w, as defined in formula (3.3). The correlation in heavy traffic with multi-server nodes can be approximated by the correlation for single server nodes, but by taking β_{F_k} / c_k as the mean service time. This can be explained as follows: in heavy traffic the system is always busy, that is why a customer always has to wait. If the system is always busy, so all servers are occupied, this means that we can simply divide the mean service time by the number of servers to get the mean service time per server. The probability that the system is in light traffic is equal to $1\text{-}p_w$. If customers do not have to wait, we can say that the correlation between the different visits to an FCFS node is almost zero.

So we assume that the correlation between two successive sojourn times of a customer exists of the waiting probability times of some expression for the correlation plus the non-waiting probability times zero. In a single-server FCFS node the waiting probability is equal to the load. So we approximate the correlation between $S_{F_k}^i$ and $S_{F_k}^{i+l}$ as follows

$$Corr\left\{S_{F_k}^i, S_{F_k}^{i+l}\right\} = p_{w_k} \frac{Corr\left\{S_{F_k}^i, S_{F_k}^{i+l}\right\}|_{M/M/1\, with\, c\mu}}{\rho_{F_k}} + (1-p_{w_k})0$$

$$= \frac{p_{w_k}}{\rho_{F_k}} Corr\left\{S_{F_k}^i, S_{F_k}^{i+l}\right\}|_{M/M/1\, with\, c\mu} \qquad \text{(B19)}$$

The covariance between two successive sojourn times S_F^i and S_F^{i+k} of a customer when this customer is fed back at least n times is given in [13]

$$Cov\left\{S_{F_k}^i, S_{F_k}^{i+l}\right\} = \frac{\rho_{F_k}(\rho_{F_k}(1-q_k)+q_k)^{l-1}\beta_{F_k}^{\ 2}}{(1-\rho_{F_k})^2}. \qquad \text{(B20)}$$

Hence we can determine the correlation between two successive sojourn times S_F^i and S_F^{i+k} of a customer when this customer is fed back at least n times

$$Corr\left\{S_{F_k}^i, S_{F_k}^{i+l}\right\} = \frac{Cov\left\{S_{F_k}^i, S_{F_k}^{i+l}\right\}}{\sqrt{Var\left\{S_{F_k}^i\right\}}\sqrt{Var\left\{S_{F_k}^{i+l}\right\}}}$$

$$= \frac{\rho_{F_k}(\rho_{F_k}(1-q_k)+q_k)^{l-1}\beta_{F_k}^{\ 2}}{(1-\rho_{F_k})^2}\left(\frac{1-\rho_{F_k}}{\beta_{F_k}}\right)^2 = \rho_{F_k}(\rho_{F_k}(1-q_k)+q_k)^{l-1}. \qquad \text{(B21)}$$

Then we can determine an approximation for the covariation between $S_{F_k}^i$ and $S_{F_k}^{i+k}$ by filling in the right expressions for the load

$$Cov\left\{S_{F_k}^i, S_{F_k}^{i+l}\right\} = Corr\left\{S_{F_k}^i, S_{F_k}^{i+l}\right\}\sqrt{Var\left\{S_{F_k}^i\right\}}\sqrt{Var\left\{S_{F_k}^{i+l}\right\}}$$

$$= \frac{p_{w_k}}{\rho_{F_k}} Corr\left\{S_{F_k}^i, S_{F_k}^{i+l}\right\}|_{M/M/1\, with\, c\mu}\ Var\left\{S_{F_k}^i\right\} \qquad \text{(B22)}$$

$$= p_{w_k}\left(\rho_{F_k}(1-q_k)+q_k\right)^{l-1}\left(\beta_{Fk}^{\ 2} + p_{w_k}\left(2-p_{w_k}\right)\frac{\beta_{Fk}^{\ 2}}{c_k^{\ 2}(1-\rho_{Fk})^2}\right).$$

Then by substitution it follows that

$$\sum_{n_k=0}^{\infty}\sum_{i\neq j} Cov\{S_{F_k}^i, S_{F_k}^j\}(1-q_k)q_k^{\ n_k}$$

$$= \sum_{n_k=0}^{\infty}\sum_{l=1}^{n-1} 2(n_k-l)p_{w_k}(\rho_{F_k}(1-q_k)+q_k)^{l-1}\left(\beta_{F_k}^{\ 2} + p_{w_k}\left(2-p_{w_k}\right)\frac{\beta_{F_k}^{\ 2}}{(1-\rho_{F_k})^2 c_k^{\ 2}}\right)(1-q_k)q_k^{\ n_k} =$$

$$= 2p_{w_k}\left(1-q_k\right)\left(\frac{(1-\rho_{F_k})^2 c_k^{\ 2}\beta_{F_k}^{\ 2} + p_{w_k}(2-p_{w_k})\beta_{F_k}^{\ 2}}{(1-\rho_{F_k})^2 c_k^{\ 2}}\right)\sum_{l=1}^{\infty}(\rho_{F_k}(1-q_k)+q_k)^{l-1}\sum_{n_k=l}^{\infty}(n_k-l)q_k^{\ n_k}$$

$$= 2\frac{p_{w_k}(1-q_k)\beta_{F_k}^{\ 2}}{(1-\rho_{F_k})^2 c_k^{\ 2}}\left((1-\rho_{F_k})^2 c_k^{\ 2} + p_{w_k}(2-p_{w_k})\right)\sum_{l=1}^{\infty}(\rho_{F_k}(1-q_k)+q_k)^{l-1}\frac{q_k^{\ l+1}}{(1-q_k)^2}$$

$$
\begin{aligned}
&= \frac{2p_{w_k}\beta_{F_k}{}^2}{(1-\rho_{F_k})^2 c_k{}^2}\left((1-\rho_{F_k})^2 c_k{}^2 + p_{w_k}(2-p_{w_k})\right)\frac{q_k{}^2(1-q_k)}{(1-q_k)^2}\frac{1}{1-q_k(\rho_{F_k}(1-q_k)+q_k)} \\
&= \frac{q_k{}^2}{(1-q_k)}\frac{2p_{w_k}\beta_{F_k}{}^2}{(1-\rho_{F_k})^2 c_k{}^2}\frac{(1-\rho_{F_k})^2 c_k{}^2 + p_{w_k}(2-p_{w_k})}{1-q_k\rho_{F_k}+q_k{}^2\rho_{F_k}-q_k{}^2} \\
&= \frac{2q_k{}^2 p_{w_k}\beta_{F_k}{}^2((1-\rho_{F_k})^2 c_k{}^2 + p_{w_k}(2-p_{w_k}))}{(1-q_k)^2(1-\rho_{F_k})^2(1-q_k\rho_{F_k}+q_k)c_k{}^2}.
\end{aligned}
\tag{B23}
$$

From the third approximation assumption, it follows that

$$Cov\{S_{PS}^i, S_{F_k}^j\} = 0. \tag{B24}$$

From the fourth approximation assumption, it follows that

$$Cov\{S_{F_k}^i, S_{F_m}^i\} = 0. \tag{B25}$$

Then finally we obtain the following approximation for the variance of the total sojourn time in the given network

$$
\begin{aligned}
Var\{S\} = &\sum_{n=0}^{\infty}(n+1)Var\{S_{PS}^1\}(1-p)p^n + \sum_{n=0}^{\infty}\sum_{i\neq j}Cov\{S_{PS}^i, S_{PS}^j\}(1-p)p^n + \\
&\sum_{k=1}^{M}\sum_{n_k=0}^{\infty} n_k Var\{S_{F_k}{}^1\}(1-q_k)q_k{}^{n_k} + \sum_{k=1}^{M}\sum_{n_k=0}^{\infty}\sum_{i\neq j}Cov\{S_{F_k}^i, S_{F_k}^j\}(1-q_k)q_k{}^{n_k} + \\
&2\sum_{k=1}^{M}\sum_{n_1=0}^{\infty}\cdots\sum_{n_M=0}^{\infty}Cov\left\{\sum_{i=1}^{n+1}S_{PS}^i, \sum_{j=1}^{n_k}S_{F_k}^j\right\}f(n_1,\ldots,n_M) + \\
&\sum_{k\neq m}\sum_{n_1=0}^{\infty}\cdots\sum_{n_M=0}^{\infty}Cov\left\{\sum_{j=1}^{n_k}S_{F_k}^i, \sum_{j=1}^{n_m}S_{F_m}^i\right\}f(n_1,\ldots,n_M) + \\
&Var\left\{\sum_{k=1}^{M}N_k\left(\{S_{PS}^1\}+E\{S_{F_k}^1\}\right)\right\} \\
= &\frac{1}{1-p}\frac{2+\rho_{PS}}{2-\rho_{PS}}\left(\frac{\beta_{PS}}{1-\rho_{PS}}\right)^2 + \frac{2p\rho_{PS}\beta_{PS}{}^2}{(1-\rho_{PS})^2(1-p)^2(2-\rho_{PS}-p+p\rho_{PS})(2-\rho_{PS})} \\
&+\sum_{k=1}^{M}\frac{q_k}{1-q_k}\left(\beta_{Fk}{}^2 + \frac{p_w(2-p_w)\beta_{Fk}{}^2}{c_k{}^2(1-\rho_{Fk})^2}\right) + \sum_{k=1}^{M}\frac{2q_k{}^2 p_{w_k}\beta_{F_k}{}^2((1-\rho_{F_k})^2 c_k{}^2 + p_{w_k}(2-p_{w_k}))}{(1-q_k)^2(1-\rho_{F_k})^2(1-q_k\rho_{F_k}+q_k)c_k{}^2} \\
&+\sum_{k=1}^{M}\frac{q_k}{(1-q_k)^2}\left(E\{S_{PS}^1\}+E\{S_{F_k}^1\}\right)^2 + \sum_{k\neq m}\frac{p_k p_m}{(1-p)^2}\left(E\{S_{PS}^1\}+E\{S_{F_k}^1\}\right)\left(E\{S_{PS}^1\}+E\{S_{F_m}^1\}\right).
\end{aligned}
\tag{B26}
$$

Appendix C

Pictures belonging to subsection 4.3.2

In the pictures below the relation between the squared coefficient of the arrival process and the squared coefficient of the departure process is shown. Every picture corresponds with a fixed load at the PS node. For different SCV's of the arrival process (X-axis) we measured the SCV's of the departure process (Y-axis). For every probability distribution and for every load the relation between the both SCV's was very close to $c_d^2 \approx \rho_{PS}^2 + (1-\rho_{PS}^2)c_a^2$. This equation is given in the pictures below to compare it with the simulated relations.

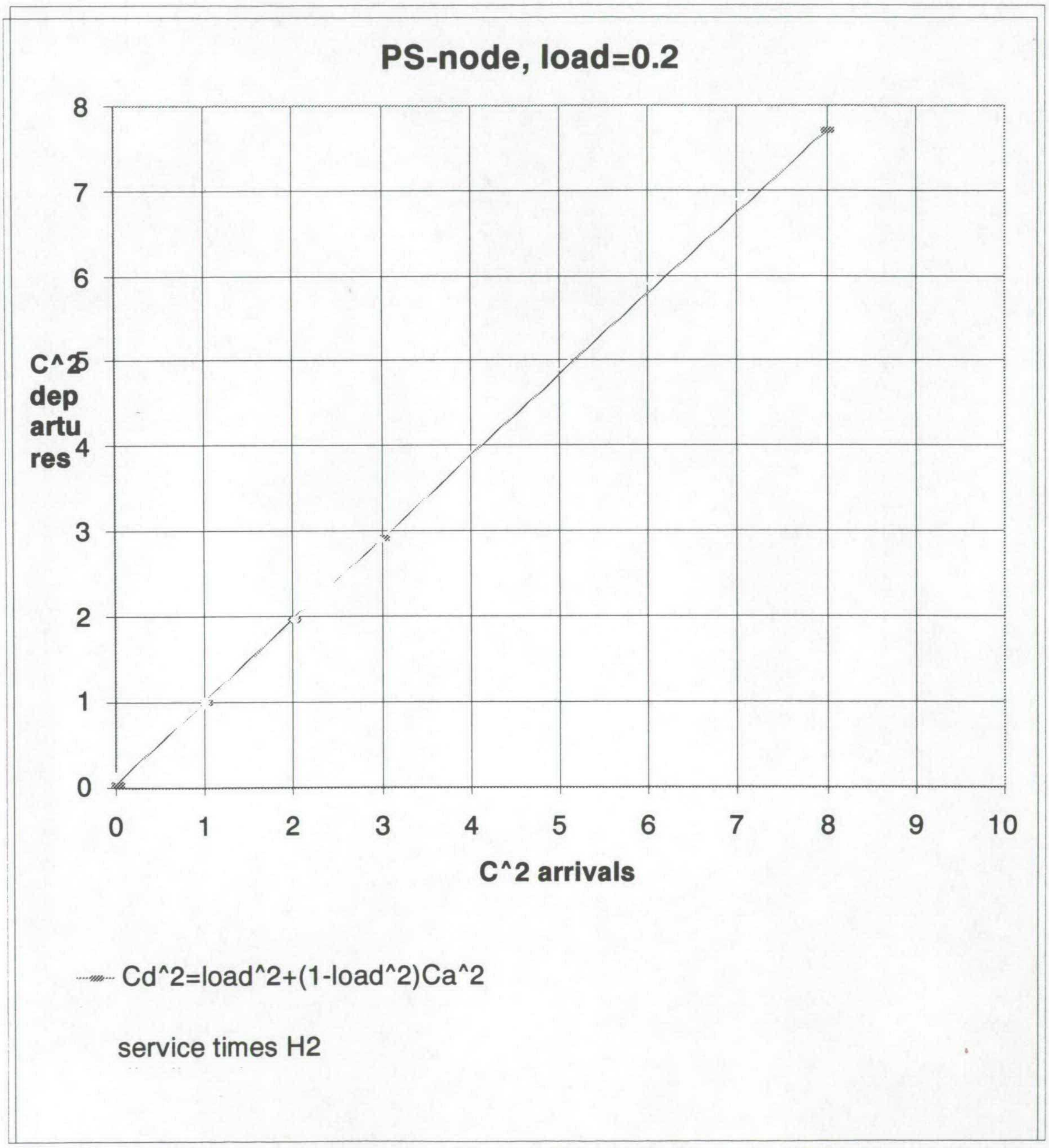


Figure B1 SCV's of the departure process at a PS node with load 0.2 related to the SCV's of the arrival process

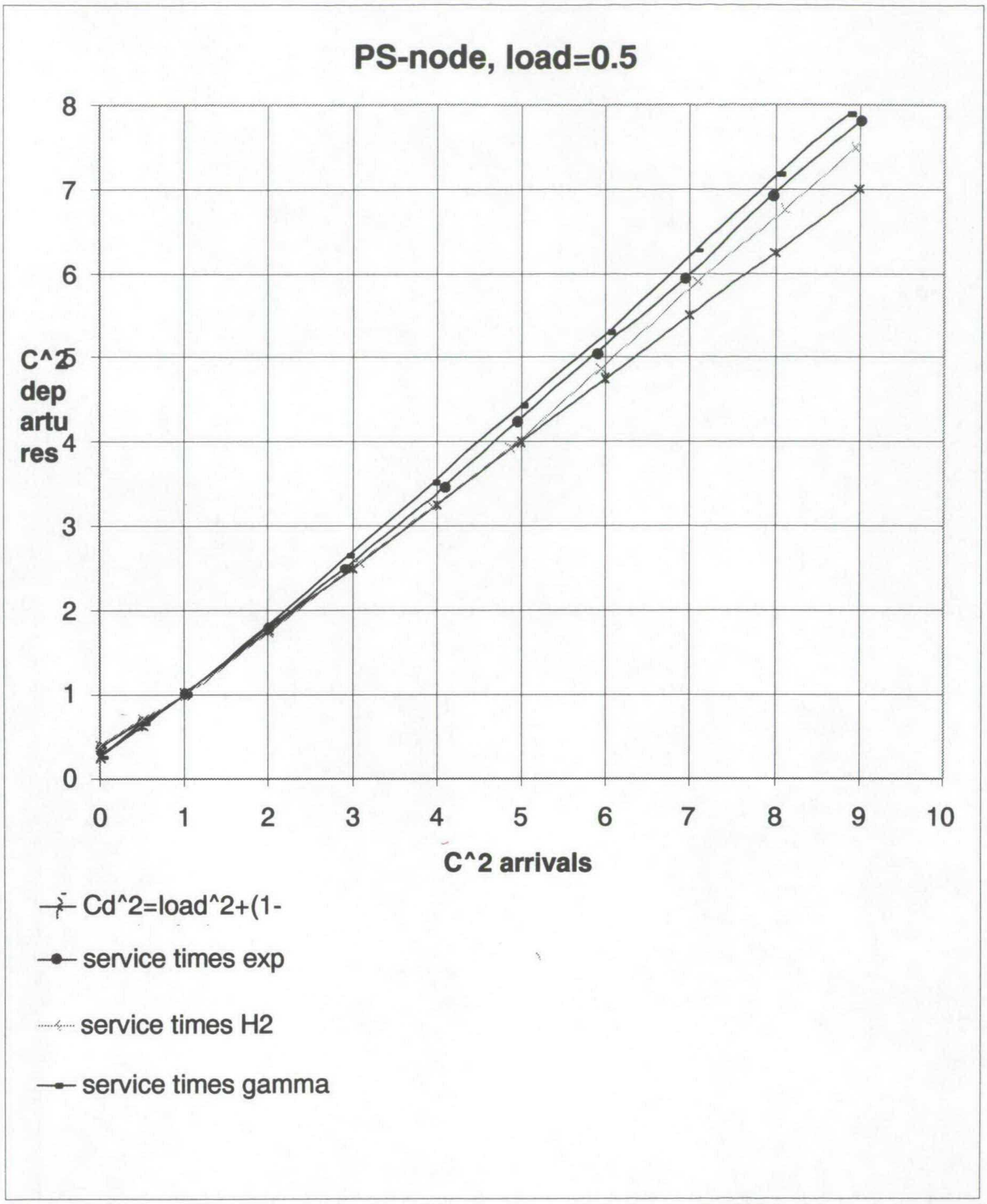


Figure B2 SCV's of the departure process at a PS node with load 0.5 related to the SCV's of the arrival process

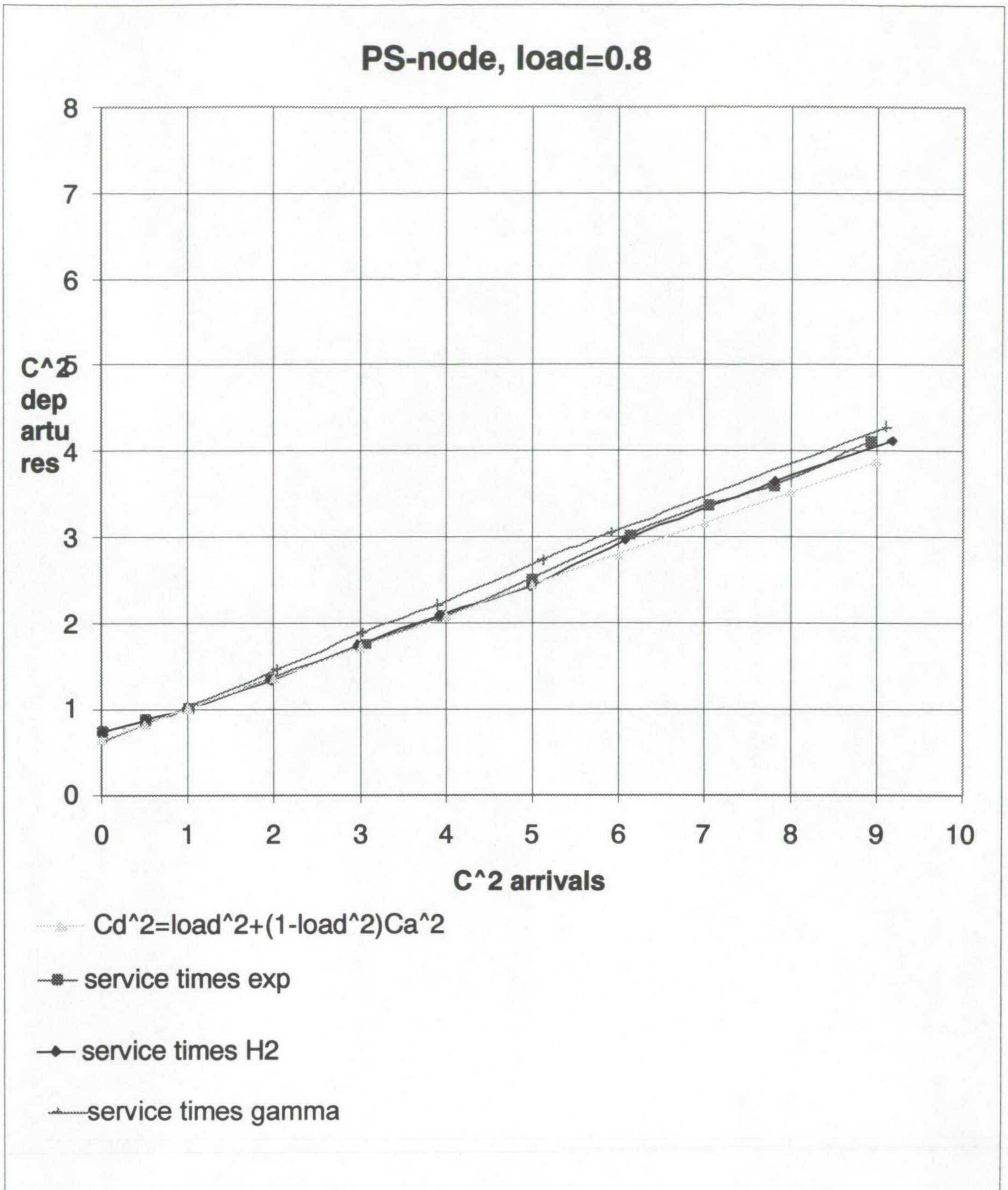


Figure B3 SCV's of the departure process at a PS node with load 0.8 related to the SCV's of the arrival process

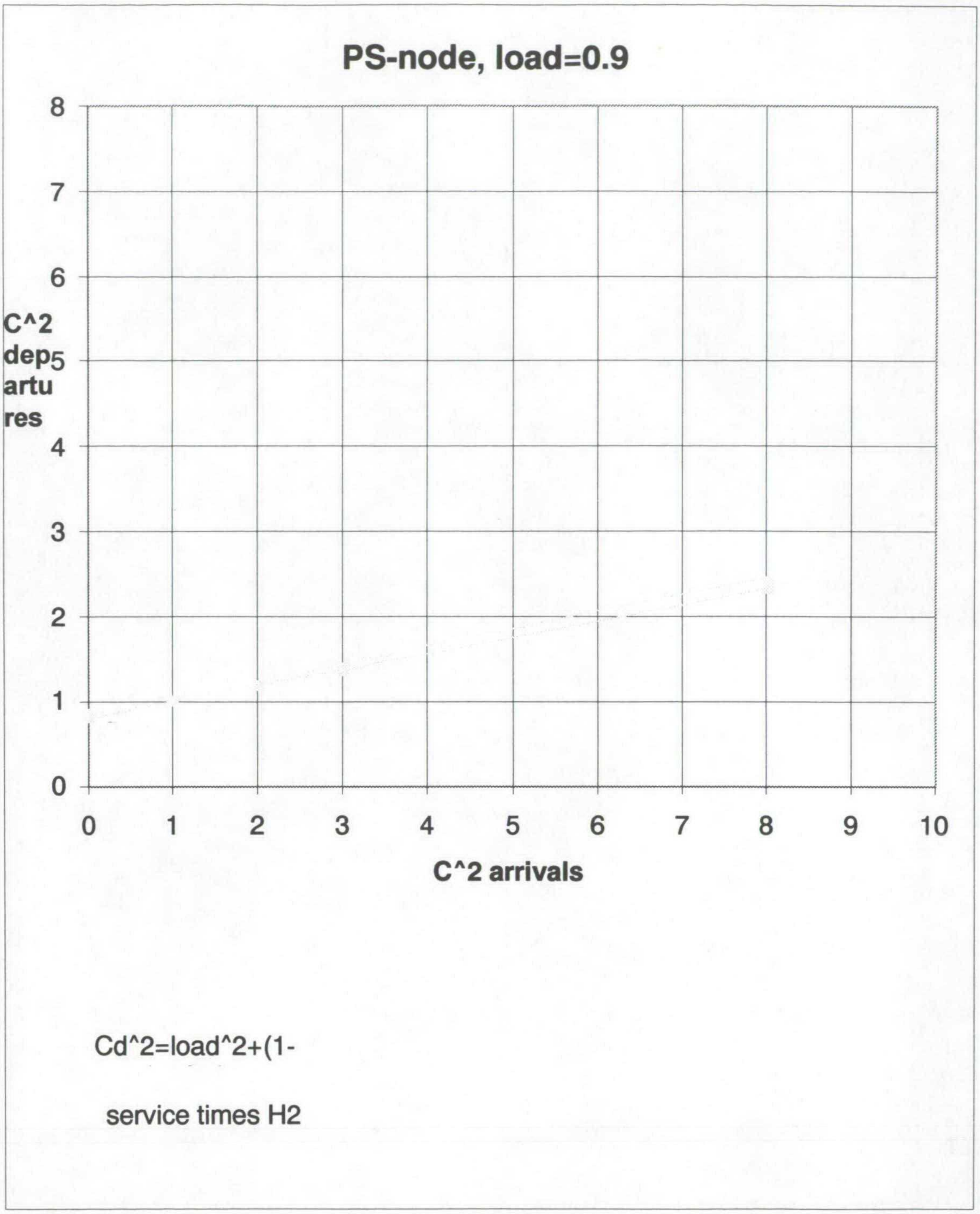


Figure B4 SCV's of the departure process at a PS node with load 0.9 related to the SCV's of the arrival process

Appendix D

Supplement to section 5.3

In a previous version of Boxma et al. [8] we found a few mistakes:

- The load at the FCFS node is said to be $p\rho_F = \frac{p\lambda\beta_F}{1-p}$. This has to be $\rho_F = \frac{p\lambda\beta_F}{1-p}$.

- The authors write that: It now follows from formula (35) of Takács [25] that:

$$E\{S_F\} \approx p\sum_{k=1}^{\infty}(1-p)p^{k-1}E\{\sigma_F^{(k)}\} = \frac{p}{1-p}\left(\frac{\beta_F}{1-\rho_F} + \frac{\lambda^2}{2}(\beta_F^{(2)} - 2\beta_F^{\,2})\frac{p}{1-p}\frac{\beta_F}{1-\rho_F}\right).$$

instead of

$$E\{S_F\} = p\sum_{k=1}^{\infty}(1-p)p^{k-1}E\{\sigma_F^{(k)}\} = \frac{p}{1-p}\left(\frac{\beta_F}{1-\rho_F} + \frac{\lambda}{2}(\beta_F^{(2)} - 2\beta_F^{\,2})\frac{p}{1-\rho_F}\right).$$

Probably the authors made more than one mistake by deriving this formula. We think that they used the wrong definition of the load at the FCFS-node given above. And probably they forgot to replace Takács' λ by $p\lambda$.

- Furthermore the authors write that: "According to Van den Berg [4], formulas (2.64) and (2.65):

$$E\{\sigma_{F,1}\} = \frac{\beta_F}{1-\rho_F} + \frac{\lambda}{2}(\beta_F^{(2)} - 2\beta_F^{\,2})\frac{1}{1-p}\frac{1-p\rho_F}{1-\rho_F},$$

and for $j = 2,3,\ldots$:

$$E\{\sigma_{F,j}\} = \frac{\beta_F}{1-\rho_F} + \frac{\lambda}{2}(\beta_F^{(2)} - 2\beta_F^{\,2})\rho_F\frac{1+p-p\rho_F}{1-\rho_F}(\lambda\beta_F + p)^{j-2}."$$

Here $E\{\sigma_{F,j}\}$ is the mean sojourn time of the j^{th} visit of a customer to the M/G/1-FCFS queue with instantaneous Bernoulli feedback.

The authors forgot to replace Van den Berg's λ by $p\lambda$.

Now we check the following relations:

$$E\{\sigma_F\} = p\sum_{k=1}^{\infty}(1-p)p^{k-1}E\{\sigma_F{}^{(k)}\} = p\sum_{k=1}^{\infty}(1-p)p^{k-1}\sum_{j=1}^{k}E\{\sigma_{F,k}\}$$

$$= pE\{\sigma_{F,1}\} + \sum_{k=1}^{\infty}(1-p)p^{k-1}\sum_{j=1}^{k}E\{\sigma_{F,k}\}$$

by substituting Van den Berg's [4] formulas for $E\{\sigma_{F,1}\}$ and $E\{\sigma_{F,k}\}$:

$$\mathrm{E}\left\{\sigma_F{}^{(1)}\right\} = \frac{\beta_F}{1-\rho_F} + \frac{p\lambda}{2}(\beta_F{}^{(2)} - 2\beta_F{}^2)\frac{1}{1-p}\frac{1-p\rho_F}{1-\rho_F},$$

$$\mathrm{E}\left\{\sigma_F{}^{(k)}\right\} = \frac{\beta_F}{1-\rho_F} + \frac{p\lambda}{2}(\beta_F{}^{(2)} - 2\beta_F{}^2)\rho_F\frac{1+p-p\rho_F}{1-\rho_F}(p\lambda\beta + p)^{k-2}, k = 2,3,....$$

We can show that calculating the sojourn time in an M/G/1-FCFS queue per customer by summing the sojourn times for every visit by Van den Berg's equations gives Takács' [25] equation.

$$E\{S_F\} \approx p\sum_{k=1}^{\infty}(1-p)p^{k-1}E\left\{\sigma_F{}^{(k)}\right\} = p\sum_{k=1}^{\infty}(1-p)p^{k-1}E\left\{\sum_{j=1}^{k}\sigma_{F,j}\right\}$$

$$= p\sum_{k=1}^{\infty}(1-p)p^{k-1}\left(E\left\{\sigma_{F,1}\right\} + \sum_{j=2}^{k}E\left\{\sigma_{F,j}\right\}\right)$$

$$= p\sum_{k=1}^{\infty}(1-p)p^{k-1}\left(\frac{\beta_F}{1-\rho_F} + \frac{p\lambda}{2}\left(\beta_F{}^{(2)} - 2\beta_F{}^2\right)\frac{1}{1-p}\frac{1-p\rho_F}{1-\rho_F} + \right.$$

$$\left.\sum_{j=2}^{k}\left(\frac{\beta_F}{1-\rho_F} + \frac{p\lambda}{2}\left(\beta_F{}^{(2)} - 2\beta_F{}^2\rho_F\right)\frac{1+p-p\rho_F}{1-\rho_F}\left(p\lambda\beta + p\right)^{j-2}\right)\right)$$

$$= p\sum_{k=1}^{\infty}(1-p)p^{k-1}\frac{\beta_F}{1-\rho_F}$$

$$+ p\sum_{k=1}^{\infty}\frac{\lambda}{2}\left(\beta_F{}^{(2)} - 2\beta_F{}^2\right)\frac{p}{1-p}\frac{1-p\rho_F}{1-\rho_F}$$

$$\sum_{j=2}^{k}\left(\frac{\beta_F}{1-\rho_F} + \frac{\lambda}{2}\left(\beta_F{}^{(2)} - 2\beta_F{}^2\right)p\rho_F\frac{1+p-p\rho_F}{1-\rho_F}\left(p\lambda\beta + p\right)^{j-2}\right)$$

$$= \frac{p}{1-p}\frac{\beta_F}{1-\rho_F} + p\sum_{k=1}^{\infty}(1-p)p^k\frac{\lambda}{2}(\beta_F{}^{(2)} - 2\beta_F{}^2)$$

$$\left(\frac{p}{1-p}\frac{1-p\rho_F}{1-\rho_F} + p\rho_F\frac{1+p-p\rho_F}{1-\rho_F}\frac{1-(p\lambda\beta+p)^{k-1}}{1-(p\lambda\beta+p)})\right)$$

$$= \frac{p}{1-p}\left(\frac{\beta_F}{1-\rho_F} + \frac{\lambda}{2}(\beta_F{}^{(2)} - 2\beta_F{}^2)\left(\frac{p}{(1-\rho_F)^2} - \frac{p\rho_F}{(1-\rho_F)^2}\right)\right)$$

$$= p\frac{1}{1-p}\left(\frac{\beta_F}{1-\rho_F} + \frac{\lambda}{2}(\beta_F{}^{(2)} - 2\beta_F{}^2)\frac{p}{1-\rho_F}\right).$$

Now, if we use Whitt's [28] formula for the mean waiting time in a GI/G/1-FCFS queue for obtaining the mean sojourn time in an M/G/1-FCFS queue, we can show that this gives the same result as Takács too:

$$E\{S_F\} = \frac{\beta_F}{1-p}\frac{\rho_F}{1-\rho_F}\frac{1+p+(1-p)c_{b_F}^{\ 2}}{2} + \frac{\beta_F}{1-p}$$

$$= \frac{\beta_F}{1-p}\frac{p\beta_F\lambda}{1-p}\frac{1}{1-\rho_F}\frac{1+p+(1-p)\left(\frac{\beta_F^{(2)}}{\beta_F^{\ 2}}-1\right)}{2} + \frac{\beta_F}{1-\rho_F}\frac{1-\rho_F}{1-p}$$

$$= \frac{1}{1-p}\left(\frac{\lambda}{2}\frac{\beta_F^{\ 2}}{1-\rho_F}\frac{p}{1-p}\left((1-p)\frac{\beta_F^{(2)}}{\beta_F^{\ 2}}+2p\right) - \frac{\rho_F\beta_F}{1-\rho_F} + \frac{\beta_F}{1-\rho_F}\right)$$

$$= \frac{1}{1-p}\left(\frac{\lambda}{2}\frac{p}{1-\rho_F}\left(\beta_F^{(2)} + 2\frac{p}{1-p}\beta_F^{\ 2}\right) - \frac{p\beta_F\lambda}{1-p}\frac{\beta_F}{1-\rho_F} + \frac{\beta_F}{1-\rho_F}\right)$$

$$= \frac{1}{1-p}\left(\frac{\lambda}{2}\frac{p}{1-\rho_F}\left(\beta_F^{(2)} + 2\frac{p}{1-p}\beta_F^{\ 2}\right) - \frac{\lambda}{2}\frac{p}{1-\rho_F}\frac{2\beta_F^{\ 2}}{1-p} + \frac{\beta_F}{1-\rho_F}\right)$$

$$= \frac{1}{1-p}\left(\frac{\lambda}{2}\frac{p}{1-\rho_F}\left(\beta_F^{(2)} + 2\frac{p}{1-p}\beta_F^{\ 2} - \frac{2\beta_F^{\ 2}}{1-p}\right) + \frac{\beta_F}{1-\rho_F}\right)$$

$$= \frac{1}{1-p}\left(\frac{\beta_F}{1-\rho_F} + \frac{\lambda}{2}\left(\beta_F^{(2)} - 2\beta_F^{\ 2}\right)\frac{p}{1-\rho_F}\right).$$

In Blanc [5] a formula for the mean waiting time of a customer in the M/G/1 system with the FCFS service discipline is given in formula (6.26). This equation is also known as the Pollaczek-Khintchine formula. By summing the mean waiting time and the mean service time, we get an equation for the mean sojourn time in an M/G/1-FCFS node

$$E\left\{S_F^{\ i}\right\} = \frac{\rho_F}{1-\rho_F}\frac{\beta_F^{(2)}}{2\beta_F} + \beta_F = \frac{\rho_F}{(1-\rho_F)}\frac{\beta_F^{\ 2}}{2}\left(1+c_{b_F}^{\ 2}\right) + \beta_F.$$

Note that this holds for M/G/1-FCFS nodes without feedback.

We can rewrite this as follows

$$E\left\{S_F^{\ i}\right\} = \frac{\rho_F}{1-\rho_F}\frac{\beta_F^{(2)}}{2\beta_F} + \beta_F = \frac{p\beta_F\lambda}{1-p}\frac{1}{1-\rho_F}\frac{\beta_F^{(2)}}{2\beta_F} + \beta_F = \frac{p}{1-p}\frac{\lambda}{2}\frac{\beta_F^{(2)}}{1-\rho_F} + \frac{\beta_F}{1-\rho_F} - \frac{\beta_F\rho_F}{1-\rho_F}$$

$$= \frac{p}{1-p}\frac{\lambda}{2}\frac{\beta_F^{(2)}}{1-\rho_F} - \frac{\beta_F}{1-\rho_F}\frac{p\beta_F\lambda}{1-p} + \frac{\beta_F}{1-\rho_F} = \frac{p}{1-p}\frac{\lambda}{2}\left(\frac{\beta_F^{(2)}}{1-\rho_F} - \frac{2\beta_F^{\ 2}}{1-\rho_F}\right) + \frac{\beta_F}{1-\rho_F} =$$

$$= \frac{\beta_F}{1-\rho_F} + \frac{\lambda}{2}\left(\beta_F^{(2)} - 2\beta_F^{\ 2}\right)\frac{p}{1-p}\frac{1}{1-\rho_F}.$$

This is not exactly the same equation as Takács' equation, because this equation does not hold for feedback, while Takács' equation holds for immediate feedback.