Product-Form Results for Two-Station Networks with Shared Resources

W. van der Weij [a], N.M. van Dijk [b], R.D. van der Mei [a,c]

[a] *CWI, Advanced Communication Networks, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands*

[b] *UVA, University of Amsterdam, Department of Econometrics, Roeterstraat 11, 1018 WB Amsterdam, The Netherlands*

[c] *VU University Amsterdam, Faculty of Sciences, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands*

Abstract

Queueing networks are studied with two stations: either in tandem or in parallel, and with a common service resource shared among the two stations. First, a necessary and sufficient criterion, called adjoint reversibility, is provided to decide whether the system possesses a product form or not. This criterion unifies both the parallel (a reversible) and the tandem (a non-reversible) system in one product-form theorem. Next, the criterion is applied separately for the parallel and tandem system to obtain a number of new product-form examples which also includes non-balanced capacity sharing. Despite of, but also due to, the different parallel and tandem mechanisms we observe that for certain examples the product form has the same structure, while for others there are essential differences. In addition, it is also proven that several models can not have a product-form result. The results provide new insights and a step forward in understanding the behavior of multi-layered queueing networks in which resources are shared among stations.

Key words: Layered queueing networks, limited processor sharing, product forms, adjoint reversibility.

1 Introduction

Over the past few decades queueing theory has been successfully applied to solve performance problems in a wide variety of application areas. A common assumption in most classical queueing models is that the servers are independent, non-interacting entities that can serve incoming jobs at a fixed rate. However, in several modern application areas, such as computer-communication

systems, the development of performance models naturally leads to the formulation of queueing networks where the servers effectively share common resources. In this type of models, the service rate at each station generally depends on the state of the entire system. Today, despite the wide applicability of queueing networks with shared resources, remarkably little is known about their behavior. Motivated by this, in this paper we study perhaps the simplest non-trivial class of queueing models in which resources are shared: a two-station network of queues, either in tandem or in parallel, in which a common resource is shared amongst the servers at both stations. For this class of models, we derive a variety of product-form and non-product-form results, leading to fundamental insight and understanding in the behavior of queueing networks with shared resources.

Queueing networks with shared resources occur naturally in the modeling of information and communication infrastructures, in which we observe a growing diversity in distributed services in which different applications share parts of the available infrastructure. In such environments, different applications compete for access to shared resources, both at the *software* layer (e.g., mutex and database locks, thread pools) and at the *hardware* layer (e.g., bandwidth, processing power, disk access). To handle incoming requests, application servers usually implement a number of thread pools, each of which is dedicated to perform a specific sub-transaction. A Web server is an example of such an application server. A particular feature of the Web server model proposed in [16,31] is that at any moment in time the active (i.e., non-idling) threads share a common CPU hardware in a processor sharing (PS) fashion. Other examples of models in which software resources compete for access to shared hardware resources are presented in [17,32]. In fact, due to the sharing of resources the actual service rates of the applications competing for this resource are dependent. An interesting line of research in which the service rates among different network stations are also dependent is focused on bandwidth-sharing networks [30,4], providing a natural modeling framework for describing the dynamic flow-level interaction among elastic data transfers in communication networks. Queueing models where resources are shared among the different stations also occur naturally in the modeling of the flow-level performance in wireline data networks where the capacity of different links are shared among competing flows [3], or in wireless networks, where a limited amount of bandwidth is shared among different users, and where users can communicate via a cascade of intermediate hops [8].

In the literature, a variety of papers focuses on queueing networks with a layered structure. In [37], Rolia and Sevcik propose the Method of Layers (MoL), i.e., a closed queueing-network model based on the responsiveness of client-server applications. Woodside et al. [41] propose the so-called Stochastic Rendez-Vous Network (SRVN) model to analyze the performance of application software with client-server synchronization. Ramesh and Perros [36]

model a Web server system where the servers form a multi-tiered structure, and where clients and servers communicate via synchronous and asynchronous communication; they propose an approximate method for calculating the mean response time based on a decomposition approach. Dilley et al. [13] describe custom instrumentation to collect workload metrics and model parameters from large-scale Web servers, and they develop a Layered Queueing Model (LQM) of a Web server and use this model to predict the impact of a single Web server thread pool size on the server and client response times. Franks et al. [15] focus on the detection of bottlenecks in the context of LQMs. Another interesting class of models in which the service rates at the different stations are dependent are the so-called coupled-processor models, i.e., multi-server models where the speed of a server at a queue depends on the number of servers at the other queues (see for example [29,9,14]). A variety of papers are focused on the so-called Limited Processor Sharing (LPS) model, a PS model in which a newly incoming job is only accepted when the number of jobs in the system is less than some threshold T; customers that find the system full are placed in an infinite-entrance buffer which is served on an First Come First Served (FCFS) basis. For the LPS model, Avi-Itzhak and Halfin [1] give a simple approximation for the expected sojourn time. Very recently, several new results for the LPS queue have been obtained. Nuyens and Van der Weij [34] derive stochastic monotonicity results of the sojourn time distribution with respect to the admittance threshold T. Zhang et al. [42,43] investigate the LPS queue, and describe the behavior of the queue in heavy traffic and derive an approximation for the waiting probability. And in [44], Zhang and Zwart derive an approximation for the steady-state queue length and response time in heavy traffic. Van der Weij [39] proposes simple approximations for the expected sojourn times for a tandem of queues with processor-shared resources. A considerable amount of research has been dedicated to the stability of layered queueing networks. Borst et al. [6] give a sharp characterization of per-station stability for parallel stations with a decreasing service allocation. Jonckheere et al. [26] derive more general results for the rate stability of networks with a general class of capacity allocation functions.

In this paper we study a class of queueing networks with a two-layered structure where the service rates of the different stations might depend on the complete system; in particular, we characterize a number of product-form as well as non-product form results. Extensive literature has appeared [24,25] providing product-form results for job-shop networks. The well known BCMP-paper for computer applications [2] and other extensions of networks having a product form can be found in [28]. Schassberger [38], Pittel [35] and Hordijk and Van Dijk [22,23] also contribute in product-form extensions, including blocking and non work-conserving service disciplines. In [20] product-form results are presented for the Coxian Processor Sharing queue with no initial blocking but mid stage exit. Specific product-form results for processor sharing systems are presented, most notably, in [7] and [5]. Most essentially thought, in

these references, the capacity allocation functions are assumed to be strictly positive. In [5] Bonald and Proutìere show that the stationary distribution of a network is insensitive for the service-time distribution if and only if the service capacities are balanced, considering networks with state dependent service rates and state dependent arrival rates. In Van Dijk [10,12] are sufficient and necessary conditions provided for a network to possess a product-form solution. The focus in these references is on blocking. In this paper, in contrast, the focus is on the sharing of the service capacity. In addition it is studied whether or not the parallel and tandem models are equivalent with respect to their product forms. In particular the product-form results are compared for the tandem and parallel model with similar sharing functions. We specify the criterion in [10,12] to give both *necessary* and *sufficient* conditions for the existence of a product-form solution to a general setting of service sharing among two stations in either parallel or tandem. A theorem is provided to unify models despite different routing mechanisms, leading to comparable (and similar) product-form solutions. The product-form behavior of a range of model examples will be analyzed. This covers the standard processor-sharing mechanism in which the resource is fairly shared among the jobs in the system; note that for this model the existence of a product form is well known, but that we give an alternative approach to prove this. Moreover, both new product-form and non-product form results for non-standard PS models are concluded, e.g., where the resource sharing may be unproportional and where service may be stopped. This analysis leads to a number of new product-form and non-product form results that have not been reported explicitly before.

The set up of this paper is as follows. In Section 2 the models investigated in this paper are described and relevant notation and definitions are introduced. Also the general condition for models to possess a product-form solution or not is given and specified. In Section 3 the parallel model is discussed in detail and examples are given for several capacity allocations and state space truncations leading to product-form solutions and non-product form solutions. In Section 4 similar results are presented for the tandem model. After a discussion in Section 5 we conclude with addressing a number of challenging topics for further research.

2 The models and general product-form characterization

We restrict the presentation to queueing networks with two service stations and investigate product-form properties of these queueing networks, where the networks have the following specific features: (1) state-dependent service sharing, where the per-station service rates depend on the state of the entire system, (2) service can be fully stopped at a station, even if jobs are present

at that station, and (3) incoming jobs can be denied access to the system. For the networks we focus on the sharing of the capacity, more then on blocking, which is motivated by the applications introduced in Section 1. We focus on networks with only two stations since the complexities with respect to product forms manifest themselves for these networks, while the behavioral insights and intuition can be obtained by illustrations.

We consider two models, both with two stations, a model with two stations in parallel (PM), and a model with two stations in tandem (TM). For these models we first introduce some common notation. Denote the state of the system by $\mathbf{n} = (n_1, n_2)$, where n_i denotes the number of jobs present (i.e., waiting or in service) at station i ($i = 1,\ 2$). The state space is denoted by C. Let the total amount of service capacity offered to all jobs in service at station i denoted by $f_i(\mathbf{n}) \geq 0$, for $i = 1,\ 2$. We assume that an empty station does not receive service capacity (i.e., $f_i(\mathbf{n}) = 0$ if $n_i = 0$). The service times at station i are exponentially distributed with mean $\beta_i = \mu_i^{-1}$. Given this notation, we now define the two different models.

2.1 Parallel Model (PM)

Consider a network of two stations in parallel, we denote this model the parallel model (PM). Jobs arrive at station i according to a Poisson process with rate λ_i ($i = 1,\ 2$). After completion of service at station i a job leaves the network. Upon arrival at station i, an incoming job is either accepted or blocked, depending on the state of the system, $\mathbf{n}$. This admission policy, denoted by the blocking function $b_i(\mathbf{n}) \in \{0, 1\}$ ($i = 1,\ 2$), is defined as follows: If $b_i(\mathbf{n}) = 1$ then a job arriving at station i is accepted, and if $b_i(\mathbf{n}) = 0$ the job is blocked. In Section 3 we focus on product forms for this model, given a function $f_i(\mathbf{n})$, for $i = 1,\ 2$. A first example of this model is presented in Figure 1. In this example, which will be discussed in detail in Section 3.5 the state space $\mathbf{n}$ equals $(4, 2)$. The capacity assignment is based on a processor sharing discipline, jobs in service receive a fair share of the total capacity, and in this example three jobs are in service in the first station, and two jobs in the second station, all receiving a fifth of the total capacity of the commonly shared resource.

2.2 Tandem Model (TM)

For the tandem model (TM) we consider a network consisting of two stations in tandem. The jobs arrive at station 1 according to a Poisson process with rate λ. After completion of service at this station jobs are forwarded to station 2; after receiving service at station 2 jobs depart from the network. There are

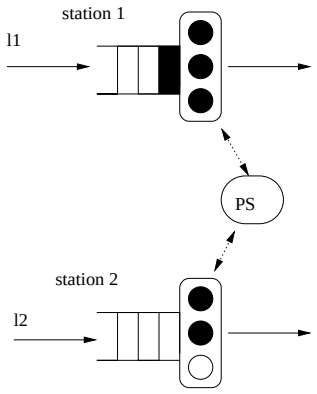

Fig. 1. The Parallel Model.

no external arrivals to the second station. Upon arrival at the system, an incoming job is either accepted or blocked, depending on the state of the system. To this end, we again denote an admission policy, for the tandem model by $b_1(\mathbf{n}), \mathbf{n} \geq \mathbf{0}$, where $b_1(\mathbf{n}) := 1$ if an arriving job is accepted, and $b_1(\mathbf{n}) := 0$ otherwise. Note that we assume that no blocking exists on station 2. In Section 4 this model and its product-form results are presented and discussed.

Figure 2 illustrates an example of this model where the capacity assignment is again based on a processor sharing discipline. Note that in this figure, as well as in Figure 1; $\mathbf{n} = (4, 2)$, with three jobs in service at the first station and two at the second station.

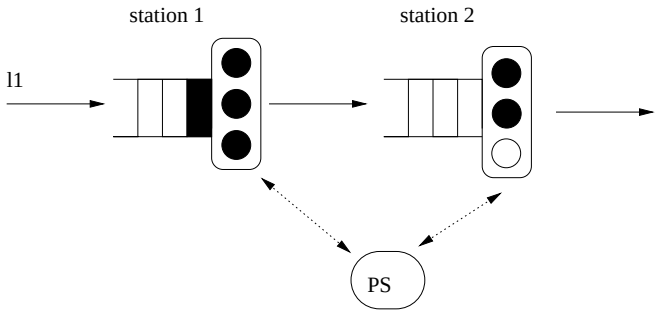

Fig. 2. The Tandem Model.

We note that the three features addressed above are included in both model descriptions. State-dependent service sharing is captured in the function $f_i(\mathbf{n})$, which includes the possibility to provide no service to station i by taking $f_i(\mathbf{n}) := 0$ for some $n_i > 0$. Access blocking is included in the definition of $b_i(\mathbf{n})$.

Under natural ergodicity assumptions for its existence, let $\pi(\mathbf{n})$ denote the corresponding steady-state distribution. In this section we present a general criterion that gives both necessary and sufficient conditions for $\pi(\mathbf{n})$ to possess a product-form solution. Here the standard perception of a product form is used in that it factorizes in structure to the stations, as specified by:

A product form is defined as the factorization of the steady-state joint station distribution to the steady-state single station distribution, up to normalization and its state space [10].

2.3.1 Station balance

As will be shown below, the existence of a product form can be characterized by the so-called notion of reversibility, not necessarily of the underlying Markov chain itself but of a special constructed Markov chain that will be called the adjoint Markov chain. This notion of reversibility reflects the phenomenon that a chain would stochastically evolve in the same way if we could reverse time (see [28] for an elegant and extensive exposure of this concept).

The construction of the adjoint Markov chain depends on the specific application of interest in order for a notion of station balance to be satisfied, i.e.

$$\begin{aligned} &\textit{The rate out of a state } \mathbf{n} \textit{ due to a departure at a station } i = \\ &\textit{the rate into that state } \mathbf{n} \textit{ due to an arrival at that station } i. \end{aligned} \tag{1}$$

Whether this station balance is indeed satisfied, which in turn appears to be directly related to a product form, then remains to be seen and is one-to-one related to the reversibility of the adjoint Markov chain (defined in Section 2.3.2). The reversibility of the adjoint Markov chain requires the existence of a stationary distribution $\bar{\pi}$, such that $\bar{\pi}(i)\bar{q}_{i,j} = \bar{\pi}(j)\bar{q}_{j,i}$, where $\bar{q}_{i,j}$ are the transition rates of the adjoint Markov chain.

Once again, it is important to observe that reversibility appears as a key characterization for a product form. However, it does *not* imply that the model *itself* needs to be reversible. Furthermore, in that case the stationary distribution $\{\pi(i)\}$ of the original chain coincides with that of the adjoint Markov chain $\{\bar{\pi}(i)\}$ up to scaling factors of the mean service times. This will be made precise by Theorem 2.1. First let us make the constructions of the ad-

joint chain explicit for the parallel and tandem model. For the parallel model the construction of the adjoint transition rates appears to be identical up to service scaling, as of the original model. For the tandem model, in contrast, the construction of the adjoint Markov chain is necessary and different as the model itself is not reversible. It is obtained by the transition rates of the original model supplemented with transition rates in the opposite direction.

From here on we adopt the state notation $\mathbf{n} = (n_1, n_2)$ as in Section 2, with n_i the number of jobs at station $i = 1,\ 2$, and we assume the existence of a stationary distribution $\pi(\mathbf{n})$ at some set of admissible states C. Hence, $\pi(\mathbf{n}) = 0$ for $\mathbf{n} \notin C$. The following notation is convenient throughout. Let $\mathbf{e}_i$ denote the i^{th} unit vector, for $i = 1,\ 2$, and let $\mathbf{0} := (0, \ldots, 0)$. Finally, denote by $\ _E$ the indicator function for an event E, i.e. $\ _E = 1$ if event E is satisfied and 0 if not. We recall Sections 2.1 and 2.2 for the model descriptions.

For the **parallel model** the Kolmogorov or global balance equations for a state $\mathbf{n} \in C$, become:

$$\begin{aligned}
&\left\{\begin{array}{ll}
\pi(\mathbf{n})\lambda_1 b_1(\mathbf{n}) + & (2.1)\\
\pi(\mathbf{n})\lambda_2 b_2(\mathbf{n}) + & (2.2)\\
\pi(\mathbf{n})\mu_1 f_1(\mathbf{n}) + & (2.3)\\
\pi(\mathbf{n})\mu_2 f_2(\mathbf{n}) & (2.4)
\end{array}\right\}\\
&= \qquad (2)\\
&\left\{\begin{array}{ll}
\pi(\mathbf{n}+\mathbf{e}_1)\mu_1 f_1(\mathbf{n}+\mathbf{e}_1) + & (2.1')\\
\pi(\mathbf{n}+\mathbf{e}_2)\mu_2 f_2(\mathbf{n}+\mathbf{e}_2) + & (2.2')\\
\pi(\mathbf{n}-\mathbf{e}_1)\lambda_1 b_1(\mathbf{n}-\mathbf{e}_1) + & (2.3')\\
\pi(\mathbf{n}-\mathbf{e}_2)\lambda_2 b_2(\mathbf{n}-\mathbf{e}_2) & (2.4')
\end{array}\right\}
\end{aligned}$$

For the **tandem model** the global balance equations are, for $\mathbf{n} \in C$:

$$\begin{aligned}
&\left\{\begin{array}{ll}
\pi(\mathbf{n})\lambda b_1(\mathbf{n}) + & (3.1)\\
\pi(\mathbf{n})\mu_1 f_1(\mathbf{n}) + & (3.2)\\
\pi(\mathbf{n})\mu_2 f_2(\mathbf{n}) & (3.3)
\end{array}\right\}\\
&= \qquad (3)\\
&\left\{\begin{array}{ll}
\pi(\mathbf{n}+\mathbf{e}_2)\mu_2 f_2(\mathbf{n}+\mathbf{e}_2) + & (3.1')\\
\pi(\mathbf{n}+\mathbf{e}_1-\mathbf{e}_2)\mu_1 f_1(\mathbf{n}+\mathbf{e}_1-\mathbf{e}_2) + & (3.2')\\
\pi(\mathbf{n}-\mathbf{e}_1)\lambda b_1(\mathbf{n}-\mathbf{e}_1) & (3.3')
\end{array}\right\}
\end{aligned}$$

We cannot expect to obtain analytic solutions for equations (2) and (3), unless these equations are satisfied by the more detailed equations $(2.i)$=$(2.i)'$ for $i = 1,\ 2,\ 3,\ 4$ for the parallel model and $(3.i)$=$(3.i)'$ for $i = 1,\ 2,\ 3$ for the tandem model. These more detailed relations will be referred to as *station balance* relations.

2.3.2 Adjoint Markov chains

In this section we will define the adjoint transition rates $\bar{q}$ for the parallel and the tandem model. For the **parallel model**, as the routing itself can be seen as reversible, the transition rates of the adjoint Markov chain can be chosen as for the original Markov chain, up to service scaling by:

$$\begin{aligned}
\bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_1) &:= \lambda_1 b_1(\mathbf{n}),\\
\bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_2) &:= \lambda_2 b_2(\mathbf{n}),\\
\bar{q}(\mathbf{n}, \mathbf{n}-\mathbf{e}_1) &:= f_1(\mathbf{n}),\\
\bar{q}(\mathbf{n}, \mathbf{n}-\mathbf{e}_2) &:= f_2(\mathbf{n}),\\
\bar{q}(\mathbf{n}_1, \mathbf{n}_2) &:= 0, \text{ otherwise.}
\end{aligned} \tag{4}$$

For the **tandem model** the routing has a triangular form and is not reversible itself, since transitions only take place in one direction. In line with the detailed equations $(3.i)$=$(3.i)'$ for $i = 1,\ 2,\ 3$, therefore, define the adjoint Markov chain by constructing transition rates in opposite direction as follows:

$$\begin{aligned}
\bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_1) &:= \lambda b_1(\mathbf{n}),\\
\bar{q}(\mathbf{n}, \mathbf{n}-\mathbf{e}_1+\mathbf{e}_2) &:= f_1(\mathbf{n}),\\
\bar{q}(\mathbf{n}, \mathbf{n}-\mathbf{e}_2) &:= f_2(\mathbf{n}),
\end{aligned}$$

supplemented with

$$\begin{aligned}
\bar{q}(\mathbf{n}+\mathbf{e}_1, \mathbf{n}) &:= f_1(\mathbf{n}-\mathbf{e}_1+\mathbf{e}_2),\\
\bar{q}(\mathbf{n}-\mathbf{e}_1+\mathbf{e}_2, \mathbf{n}) &:= f_2(\mathbf{n}-\mathbf{e}_2),\\
\bar{q}(\mathbf{n}-\mathbf{e}_2, \mathbf{n}) &:= \lambda b_1(\mathbf{n}),\\
\bar{q}(\mathbf{n}_1, \mathbf{n}_2) &:= 0, \text{ otherwise.}
\end{aligned} \tag{5}$$

In the above Equations (4) and (5) the μ_i rates are not shown due to the scaling up to multiples of the service rate in the product form result. Note furthermore that this adjoint Markov chain coincides with the parametrization of the original tandem network up to exponential service parameters in the

natural station flow direction from station i to station $i+1$. In contrast though, also a flow in opposite direction has been constructed. The general definition of the transition rates of the adjoin Markov chain are as follows. Consider a queue i and a transition rate γ from queue i to some queue $i+1$, then

$$\bar{q}(\mathbf{n}+\mathbf{e}_i, \mathbf{n}+\mathbf{e}_{i+1}) := \gamma \quad \textit{(as original Markov chain)},$$

and

$$\bar{q}(\mathbf{n}+\mathbf{e}_i, \mathbf{n}+\mathbf{e}_{i-1}) := \gamma \quad \textit{(as new)}.$$

2.3.3 Product-form result

Both the parallel and the tandem model can now be characterized by one unifying theorem. To the best of the authors knowledge, this seems to be new in the literature. It characterizes the existence of a product-form solution by means of reversibility of the adjoint Markov chain, which we will refer to as *adjoint reversibility*.

Theorem 2.1 *There exists a product-form steady-state distribution of the form*

$$\pi(\mathbf{n}) = cH(\mathbf{n}) \prod_i \left[\frac{1}{\mu_i}\right]^{n_i}, \quad \textit{for all } \mathbf{n} \in C \tag{6}$$

with c a normalizing constant, if and only if the adjoint Markov chain is reversible. That is for some steady-state distribution $H(\mathbf{n})$ and for all pairs of states $\mathbf{n}_1, \mathbf{n}_2 \in C$:

$$H(\mathbf{n}_1)\bar{q}(\mathbf{n}_1, \mathbf{n}_2) = H(\mathbf{n}_2)\bar{q}(\mathbf{n}_2, \mathbf{n}_1). \tag{7}$$

Proof: The proof is concluded directly by substitution of Equation (4) in Equation (2) and showing that $(2.i) = (2.i')$ for $i = 1,\ 2,\ 3,\ 4$ for the parallel model, and similarly, by substitution (5) in (3) showing that $(3.i) = (3.i')$ for $i = 1,\ 2,\ 3$ for the tandem model. □

When we consider the first case, start with the state (0,0). The total rate in and rate out can then be considered by showing that (2.1)=(2.1′). Next consider state (1,0) and first assume that the system does allow not more than one job, then we find the relation from state (0,0) to (0,1) and (1,0). This relation can be used when relaxing the constraint from one job in the system to two jobs and need to be filled in Equation (2). Continue this recursive method and find the proof of the theorem.

2.3.4 Reversibility characterization

The major advantage of Theorem 2.1 is that it enables one to verify the existence of a product form (6), by simply investigating the existence of a reversible solution $H(\mathbf{n})$. This in turn, can be verified by the so-called Kolmogorov criterion (see for example [33]) as based upon just the transition rates as defined by (4) and (5).

Below we present the detailed reversibility conditions in more detail, for two reasons:

1. For the readability of the paper, and
2. To use these reversibility characterizations explicitly later on in the proofs for product-form and non-product form results for the parallel and tandem model.

To verify reversibility of the adjoint Markov chain, we need to verify if one of the two conditions below, (8) or (11) holds.

Lemma 2.2 *(Equivalent adjoint reversibility conditions)*
Either of the following two conditions are equivalent for the reversibility of the adjoint Markov chain as in (7)*. The Kolmogorov equations are verified* (7) *and the product-form solution* (6) *exists, if and only if:*

1. *For any cycle of the form p of any length t and its reverse cycle of the form $\bar{p}$:*

$$\theta(p) = \theta(\bar{p}), \tag{8}$$

where,

$$\begin{aligned} p &:= \mathbf{n}_0 \to \mathbf{n}_1 \to \ldots \to \mathbf{n}_t \to \mathbf{n}_{t+1} = \mathbf{n}_0, \\ \bar{p} &:= \mathbf{n}_0 = \mathbf{n}_{t+1} \to \mathbf{n}_t \to \ldots \to \mathbf{n}_1 \to \mathbf{n}_0, \end{aligned} \tag{9}$$

with their products of transitions rates:

$$\begin{aligned} \theta(p) &:= \bar{q}(\mathbf{n}_0, \mathbf{n}_1)\bar{q}(\mathbf{n}_1, \mathbf{n}_2) \ldots \bar{q}(\mathbf{n}_t, \mathbf{n}_0), \\ \theta(\bar{p}) &:= \bar{q}(\mathbf{n}_0, \mathbf{n}_t)\bar{q}(\mathbf{n}_t, \mathbf{n}_{t-1}) \ldots \bar{q}(\mathbf{n}_1, \mathbf{n}_0). \end{aligned} \tag{10}$$

2. *There exists a function $H(\mathbf{n})$ such that for any fixed $\mathbf{n}_0 \in C$ and any state $\mathbf{n} \in C$ it holds that*

$$\begin{aligned} H(\mathbf{n}) = H(\mathbf{n}_0) \prod_{k=0}^{K-1} \left[\tfrac{\bar{q}(\mathbf{n}_k, \mathbf{n}_{k+1})}{\bar{q}(\mathbf{n}_{k+1}, \mathbf{n}_k)} \right], \text{ for any path } \mathbf{n}_0 \to \\ \ldots \to \mathbf{n}_K = \mathbf{n}, \text{ for which the denominator is positive.} \end{aligned} \tag{11}$$

This means that $H(\mathbf{n})$ is independent of the path $\mathbf{n}_1 \to \ldots \to \mathbf{n}_{K-1}$; it only depends on $\mathbf{n}_0$ and $\mathbf{n}_K$.

Proof: This can be concluded from substitution of Equation (11) in (7) or indirectly as by [28] for the characterization of reversibility. □

Either one of the two checks above in turn can generally be reduced to basic cycles or short paths that directly suggest a necessary form of the function $H(\mathbf{n})$ and a decomposition in a service and routing component, satisfying:

$$\frac{H(\mathbf{n}+\mathbf{e}_i)}{H(\mathbf{n}+\mathbf{e}_j)} = \frac{f_i(\mathbf{n}+\mathbf{e}_i)}{f_j(\mathbf{n}+\mathbf{e}_j)} \frac{b_i(\mathbf{n})}{b_j(\mathbf{n})}. \tag{12}$$

Note that for $\mathbf{n} \in C$ if $\mathbf{n}+\mathbf{e}_j \notin C$ then $b_j(\mathbf{n}) = 0$. This equation appears to be the most explicit form to find a suggestion for the function $H(\mathbf{n})$.

Remark 2.3 *From the condition given in Equation (12) it follows that the structure of the product form does not depend on the routing mechanism, whether parallel or in tandem.*

For the applications in Sections 3.5 and 4.5 also the following corollary will appear to be usefull.

Corollary 2.4 *A product form does not exist if for some pair of states* $\mathbf{n}_s$ *and* $\mathbf{n}_t$ *for some paths* p_1 *and* p_2 *and their reversed paths* $\bar{p}_1$ *and* $\bar{p}_2$*:*

$$\Theta(p_1) \neq \Theta(p_2) \tag{13}$$

with

$$\Theta(p_i) := \frac{\theta(p_i)}{\theta(\bar{p}_i)}, \tag{14}$$

and where p_1 *and* p_2 *are paths defined as follows:*

$$\begin{aligned} p_1 &:= \mathbf{n}_s \to \mathbf{n}_1 \to \ldots \to \mathbf{n}_{K-1} \to \mathbf{n}_K = \mathbf{n}_t \\ p_2 &:= \mathbf{n}_s \to \mathbf{n}'_1 \to \ldots \to \mathbf{n}'_{K'-1} \to \mathbf{n}'_{K'} = \mathbf{n}_t \end{aligned} \tag{15}$$

Remark 2.5 *(Literature) The concept of an adjoint (artificial) Markov chain to characterize the existence of a product form has first been introduced and exploited in [22] and extended in [23]. For the case of a single job-class this characterization has been explored extensively in [10]. A somewhat related product-form characterization as by an invariance condition has also been provided in [7] under the condition that there is no blocking and that the service rates are strictly positive. Its result is included by the current one as a special case. More specifically, the most closely related results for processor sharing mechanisms are those from [5] and [7]. In these references though the implicit but essential condition is assumed for the existence of a function q (see [7]) or* Φ *(in [5]) to be seen as the function* $H(\mathbf{n})$ *in Theorem 2.1. However, these are hard to find in general. The present setting, in contrast, does lead to a*

construction or check of this function by means of reversibility, as will be illustrated in Sections 3 and 4 for the models of our interest.

Remark 2.6 *(Reversed Compound Agent Theorem (RCAT)) The RCAT theorem as presented in [18,20,19] generate the reversed process and the product form. In the most general conditions in [21], an element of local state-dependency is allowed for the active shared rates in the systems. Some of the examples given in the next sections do therefore overlap with results of the RCAT papers. However, we provide a very rigorous and clear way to derive the product forms., which gives intuitive results. Furthermore, in this paper some product-forms that may not be possible with the RCAT given the state-dependency in the rate functions.*

2.3.5 Examples

In the next two sections the theorem presented in this section is used to investigate product-form results for the following six examples for both parallel and tandem models. The parallel case is given Section 3, and the tandem case in Section 4:

(1) The proportional PS-model,
(2) An unproportional PS-model with full capacity to one station,
(3) An α-unproportional PS-model,
(4) A state-space reduction,
(5) A two-station limited PS-model,
(6) A truncated two-station limited PS-model.

The first model is well known to possess a product form. However, it is included to illustrate Theorem 2.1. The results for the second and third model seem to be new in literature, unbalanced sharing of the service capacity is captured in these examples. The results for the fourth model are known for the parallel model, but new for the tandem model; it illustrates the differences that appear between tandem and parallel routing mechanisms for truncation of the state space. The fifth model example was already introduced [1,39], but the non-product form proof is new as is the product-form truncation in the tandem case (Example 6). None of the examples did appear this detailed in literature, and therefore also contributes to the insights of product-form results.

3 Parallel Model

In this section we apply Theorem 2.1 to show and prove the existence of product-form solutions for the parallel model with shared resources as de-

scribed in Section 2.3. To this end, we write,

$$f_i(\mathbf{n}) = \Phi(n_1 + n_2)s_i(\mathbf{n}), \text{ for } i = 1,\ 2, \tag{16}$$

where $\Phi(k) > 0$ represents the *total* service capacity of the shared resource when the total number of jobs n_1+n_2 equals k, and where the sharing function $s_i(\mathbf{n})$ is the fraction of this capacity allocated to station i ($i = 1,\ 2$). Note that $f_i(\mathbf{n})$ is uniquely defined by $\Phi(n_1+n_2)$ and $s_i(\mathbf{n})$ up to a scaling constant and note that in general $\Phi(\cdot)$ is not necessarily equal to 1. We now consider the examples presented in Section 2.3.5.

3.1 Example: Proportional PS-model

Consider the two-station extension of the standard single-station PS queue where the total capacity equals $\Phi(n_1 + n_2)$ and where the fraction of this capacity allocated to the stations equals:

$$s_i(\mathbf{n}) := \frac{n_i}{n_1 + n_2}, \text{ for } i = 1,\ 2, \tag{17}$$

for $\mathbf{n} \in C$, with

$$C = \{\mathbf{n} \mid n_1, n_2 \geq 0\}. \tag{18}$$

Thus, for given state $\mathbf{n}$ station i gets a fraction $s_i(\mathbf{n})$ of the capacity $\Phi(n_1+n_2)$; in words, $s_i(\mathbf{n})$ represents the proportion of jobs that are at station i. The admission policy is given by $b_i(\mathbf{n}) := 1$ for $i = 1\ 2$ and all $\mathbf{n} \in C$, i.e., all arriving jobs are accepted for all $\mathbf{n} \in C$. Note that the classical PS-case occurs as a special case by taking $\Phi(k) = 1$ for all $k \geq 0$. Furthermore, we define:

$$P(\mathbf{n}) := \left[\prod_{k=1}^{n_1+n_2} \Phi(k)\right]^{-1}, \tag{19}$$

which we will from now on use in the remainder of the paper.

Result 3.1 *The proportional parallel PS-model possesses a product-form solution of the form* (6)*, with*

$$H(\mathbf{n}) = \left[\prod_{i=1}^{2} \lambda_i^{n_i}\right] P(\mathbf{n}) \binom{n_1 + n_2}{n_1}, \text{ for } \mathbf{n} \in C. \tag{20}$$

Proof: We first use Theorem 2.1 to prove the existence of the product form, and then, to prove that the product-form solution has the form (20). To construct a proof based on Theorem 2.1, it suffices to show that the reversibility condition (8), i.e. $\theta(p) = \theta(\bar{p})$, is satisfied for each path p. To this end, note

that for the model under consideration, the transition rates are as follows: For $\mathbf{n} \in C$,

$$\begin{aligned}
\bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_1) &= \lambda_1, \\
\bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_2) &= \lambda_2, \\
\bar{q}(\mathbf{n}, \mathbf{n}-\mathbf{e}_1) &= \tfrac{n_1}{n_1+n_2}\Phi(n_1+n_2), \\
\bar{q}(\mathbf{n}, \mathbf{n}-\mathbf{e}_2) &= \tfrac{n_2}{n_1+n_2}\Phi(n_1+n_2).
\end{aligned} \tag{21}$$

Based on these transition rates one may verify that the transition matrix of the adjoint Markov chain $\bar{Q}$ equals the transition matrix Q of the original Markov chain. Note that for this model it suffices to consider only two basic cycles, since all other cycles are constructed similarly. Thus, we only need to show that $\theta(p) = \theta(\bar{p})$ for the following two paths:

$$\begin{aligned}
p &= \mathbf{n} \to \mathbf{n}+\mathbf{e}_1 \to \mathbf{n}+\mathbf{e}_1+\mathbf{e}_2 \to \mathbf{n}+\mathbf{e}_2 \to \mathbf{n}, \\
\bar{p} &= \mathbf{n} \to \mathbf{n}+\mathbf{e}_2 \to \mathbf{n}+\mathbf{e}_1+\mathbf{e}_2 \to \mathbf{n}+\mathbf{e}_1 \to \mathbf{n}.
\end{aligned} \tag{22}$$

To this end, substitution of Equations (21) into Equation (10) leads to

$$\begin{aligned}
\theta(p) &= \bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_1)\bar{q}(\mathbf{n}+\mathbf{e}_1, \mathbf{n}+\mathbf{e}_1+\mathbf{e}_2) \cdot \\
&\quad \bar{q}(\mathbf{n}+\mathbf{e}_1+\mathbf{e}_2, \mathbf{n}+\mathbf{e}_2)\bar{q}(\mathbf{n}+\mathbf{e}_2, \mathbf{n}) \\
&= \lambda_1 \cdot \lambda_2 \cdot \tfrac{(n_1+1)}{n_1+n_2+2}\Phi(n_1+n_2+2) \cdot \\
&\quad \tfrac{(n_2+1)}{n_1+n_2+1}\Phi(n_1+n_2+1),
\end{aligned}$$

and

$$\begin{aligned}
\theta(\bar{p}) &= \bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_2)\bar{q}(\mathbf{n}+\mathbf{e}_2, \mathbf{n}+\mathbf{e}_1+\mathbf{e}_2) \cdot \\
&\quad \bar{q}(\mathbf{n}+\mathbf{e}_1+\mathbf{e}_2, \mathbf{n}+\mathbf{e}_1)\bar{q}(\mathbf{n}+\mathbf{e}_1, \mathbf{n}) \\
&= \lambda_2 \cdot \lambda_1 \cdot \tfrac{(n_2+1)}{n_1+n_2+2}\Phi(n_1+n_2+2) \cdot \\
&\quad \tfrac{(n_1+1)}{n_1+n_2+1}\Phi(n_1+n_2+1),
\end{aligned}$$

which immediately implies $\theta(p) = \theta(\bar{p})$. Hence, the reversibility condition (8) applies, and thus, there exists a product-form solution (6). Next, we show that the product-form solution has the form (20). To this end, we observe that using Equation (7) in Theorem 2.1 and the equations in (21) imply the following recursive relations for $\mathbf{n} \in C$:

$$\begin{aligned}
H(\mathbf{n})\lambda_1 &= H(\mathbf{n})\bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_1) \\
&= H(\mathbf{n}+\mathbf{e}_1)\bar{q}(\mathbf{n}+\mathbf{e}_1, \mathbf{n}) \\
&= H(\mathbf{n}+\mathbf{e}_1)\tfrac{(n_1+1)}{n_1+n_2+1}\Phi(n_1+n_2+1).
\end{aligned} \tag{23}$$

Similarly by (7) and (21) we find that,

$$\begin{aligned} H(\mathbf{n})\lambda_2 &= H(\mathbf{n})\bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_2) \\ &= H(\mathbf{n}+\mathbf{e}_2)\bar{q}(\mathbf{n}+\mathbf{e}_2, \mathbf{n}) \\ &= H(\mathbf{n}+\mathbf{e}_2)\tfrac{(n_2+1)}{n_1+n_2+1}\Phi(n_1+n_2+1), \end{aligned} \tag{24}$$

and

$$\begin{aligned} H(\mathbf{n})\mu_1\tfrac{n_1}{n_1+n_2}\Phi(n_1+n_2) &= H(\mathbf{n})\bar{q}(\mathbf{n}, \mathbf{n}-\mathbf{e}_1) \\ &= H(\mathbf{n}-\mathbf{e}_1)\bar{q}(\mathbf{n}-\mathbf{e}_1, \mathbf{n}) \\ &= H(\mathbf{n}-\mathbf{e}_1)\lambda_1, \end{aligned} \tag{25}$$

$$\begin{aligned} H(\mathbf{n})\mu_2\tfrac{n_1}{n_1+n_2}\Phi(n_1+n_2) &= H(\mathbf{n})\bar{q}(\mathbf{n}, \mathbf{n}-\mathbf{e}_2) \\ &= H(\mathbf{n}-\mathbf{e}_2)\bar{q}(\mathbf{n}-\mathbf{e}_2, \mathbf{n}) \\ &= H(\mathbf{n}-\mathbf{e}_2)\lambda_2. \end{aligned} \tag{26}$$

Note that Equation (23) equals Equation (25), since for $(n_1, n_2) = (0, 0)$ transition rates to states $(n_1 - 1, n_2) = (-1, 0)$ and $(n_1, n_2 - 1) = (0, -1)$ are zero, which forces that Equation (23) and Equation (25) are equivalent. Similarly, we conclude that Equation (24) equals Equation (26). Thus, the recursive relation can be rewritten as,

$$\tfrac{H(\mathbf{n}-\mathbf{e}_1)}{H(\mathbf{n})} = \tfrac{1}{\lambda_1}\tfrac{n_1}{n_1+n_2}\Phi(n_1+n_2),\ n_1 > 0, \tag{27}$$

$$\tfrac{H(\mathbf{n}-\mathbf{e}_2)}{H(\mathbf{n})} = \tfrac{1}{\lambda_2}\tfrac{n_2}{n_1+n_2}\Phi(n_1+n_2),\ n_2 > 0. \tag{28}$$

Equation (20) can now be easily obtained by recursively solving (27) and (28), starting with $H(\mathbf{0}) := \Phi(0)$. □

Remark 3.2 *(Alternative approaches) Instead of by the proof presented above, Result 3.1 can also be concluded:*

(1) Directly by substituting (7) in (6).
(2) From [7]. To this end, consider the system as a single processor. Let a class-r job have a service with respective parameters μ_{r_i} for class r_i. This has a one to one correspondence with the two station parallel model, since each class corresponds to a station.
(3) From [5] directly for a processor sharing disciplines and indirectly for arbitrary disciplines as in this reference it is implicitly assumed that each station itself (also) has a PS-discipline. However, as the effective service rates at station 1 and 2 are independent of the service discipline in order provided the services are assumed to be exponential, in the exponential

case the product form can be concluded for arbitrary disciplines at each station.

By this reference as well as by [7] it can also be concluded that the product form is insensitive with respect to the service-time distributions.

Remark 3.3 *(Special PS-case and insensitivity) The standard type processor sharing function, that assigns an equal (fair) share $1/(n_1+n_2)$ of the total capacity $\Phi(n_1+n_2)$ to each job in service, is included by assuming that each station itself also has a PS-discipline; that, at both stations, all jobs present equally share a fraction $n_i/(n_1+n_2)$ of the total capacity. For this particular case, it can also be concluded directly from [5] or indirectly from [7] or [22,23], that the product form (6) also applies to arbitrary non-exponential service requirements with means $1/\mu_i$ at station i. This property is well known in the literature as insensitivity.*

3.2 Example: Unproportional PS-model with full capacity to one station

A first most extreme type example in which an unproportional processor sharing is effectuated is obtained by always allocating the full capacity to one station, and to fairly share this capacity among all jobs at that station. Consider the model with access blocking functions

$$\begin{aligned} b_1(\mathbf{n}) &= {}_{E_1} \text{ with } E_1 := \{\mathbf{n} \in C : n_1 = n_2 \text{ or } n_1 = n_2 + 1\}, \\ b_2(\mathbf{n}) &= {}_{E_2} \text{ with } E_2 := \{\mathbf{n} \in C : n_1 = n_2 \text{ or } n_1 = n_2 - 1\}, \end{aligned} \tag{29}$$

and with sharing functions

$$\begin{aligned} s_1(\mathbf{n}) &= {}_{E_3} \text{ with } E_3 := \{\mathbf{n} \in C : n_1 = n_2 + 1 \text{ or } n_1 = n_2 + 2\}, \\ s_2(\mathbf{n}) &= {}_{E_4} \text{ with } E_4 := \{\mathbf{n} \in C : n_1 = n_2 \text{ or } n_1 = n_2 - 1\}. \end{aligned} \tag{30}$$

The access blocking function only allows arrivals to station 1 if $n_1 = n_2$ or $n_1 = n_2 + 1$, and similarly, station-2 arrivals are accepted only if $n_1 = n_2 - 1$ or $n_1 = n_2$. The sharing function forces to assign all capacity to station 1 if $n_1 = n_2 + 1$ or $n_1 = n_2 + 2$, and to station 2 if $n_1 = n_2 - 1$ or $n_1 = n_2$. In words, if $n_1 > n_2$ then station 1 gets the full capacity, and station 2 gets the full capacity otherwise. This model will be referred to as the unproportional parallel PS-model. Using Equations (29) and (30) it is readily verified that the state space for this model is given by

$$C = \{\mathbf{n} \mid n_1 \in \{n_2 - 1, n_2, n_2 + 1, n_2 + 2\}, \text{ with } n_1, n_2 \geq 0\}. \tag{31}$$

Figure 3 illustrates the non-zero transitions at the state space of this model.

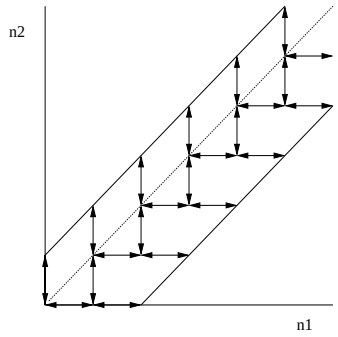

Fig. 3. Parallel Model: Transitions in the state space C for which the product form (6) applies with positive transition rates (in both directions) indicated by arrows (all other rates are equal to 0).

Result 3.4 *The unproportional parallel PS-model possesses a product-form solution of the form* (6)*, with*

$$H(\mathbf{n}) = \left[\prod_{i=1}^{2} \lambda_i^{n_i}\right] P(\mathbf{n}), \ \textit{for } \mathbf{n} \in C, \tag{32}$$

where C is defined in (31)*.*

Proof: First we show that the model possesses a product form by checking Equation (8) for all paths within the state space C, defined in (31). To this end, note that the transition rates for the adjoint Markov chain (which are again equal to the transition rates for the original Markov chain) are as follows:

$$\begin{aligned} \bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_1) &= \lambda_1, \\ \bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_2) &= \lambda_2, \\ \bar{q}(\mathbf{n}, \mathbf{n}-\mathbf{e}_1) &= \Phi(n_1+n_2), \\ \bar{q}(\mathbf{n}, \mathbf{n}-\mathbf{e}_2) &= \Phi(n_1+n_2). \end{aligned} \tag{33}$$

Note that we only need to verify Equation (8) for the following two basic cycles:

$$\begin{aligned} p_1 &= \mathbf{n} \to \mathbf{n}+\mathbf{e}_1 \to \mathbf{n}, \\ p_2 &= \mathbf{n} \to \mathbf{n}+\mathbf{e}_2 \to \mathbf{n}. \end{aligned}$$

Substitution of (33) in (10) leads to the following two equations, for $\mathbf{n} \in C$,

$$\begin{aligned} \theta(p_1) &= \bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_1)\bar{q}(\mathbf{n}+\mathbf{e}_1, \mathbf{n}) = \lambda_1 \cdot \Phi(n_1+n_2), \\ \theta(p_2) &= \bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_2)\bar{q}(\mathbf{n}+\mathbf{e}_2, \mathbf{n}) = \lambda_2 \cdot \Phi(n_1+n_2). \end{aligned}$$

Next, notice that the paths in the opposite directions, denoted by $\bar{p}_1$ and $\bar{p}_2$, are equal to the paths p_1 and p_2, respectively. Hence, for $i = 1,\ 2$ we have

$\theta(p_i) = \theta(\bar{p}_i)$, so that the reversibility condition (8) is satisfied, which implies that the model has a product-form solution. Then, to show that (32) holds, note that arguments similar to those in Result 3.1 hold and that it is easily verified that $n_i > 0$,

$$\frac{H(\mathbf{n} - \mathbf{e}_i)}{H(\mathbf{n})} = \frac{1}{\lambda_i}\Phi(n_1 + n_2), \text{ for } i = 1,\ 2, \tag{34}$$

supplemented with the starting condition $H(\mathbf{0}) := \Phi(0)$ gives $H(\mathbf{n})$ in Equation (32). Thus the steady-state distribution has the product form (6), where $H(\mathbf{n})$ is given by Equation (32). This completes the proof of the result. □

3.3 Example: α-Unproportional PS-model

Also unproportional and non-zero sharing functions over both stations might still retain the necessary invariance (11), or equivalently (8). Consider the complete state space,

$$C = \{\mathbf{n} \mid n_1, n_2 \geq 0\}, \tag{35}$$

and a sharing function $s_i(\mathbf{n})$ in which a fraction of the capacity is assigned to station 1, and a fraction of the capacity is assigned to station 2, as follows for $\mathbf{n} \in C$:

$$(s_1(\mathbf{n}), s_2(\mathbf{n})) := \begin{cases} (1-\alpha, \alpha) & \text{if } n_1 > n_2, \\ (\alpha, 1-\alpha) & \text{if } n_1 < n_2, \\ (\alpha, \alpha) & \text{if } n_1 = n_2, \end{cases} \tag{36}$$

for an arbitrary $0 < \alpha < 1/2$. The fraction of the total capacity $\Phi(n_1 + n_2)$ a station receives is dependent on the state space. The sharing function $s_i(\mathbf{n})$ partitions the state space in three regions, namely in the region where the number of jobs in the station 1 is greater than the number of jobs present at station 2 (i.e. $n_1 > n_2$), the region where the number of jobs at station 1 is smaller than the number of jobs at station 2 (i.e. $n_1 < n_2$), and the region where the number of jobs is equal in both stations (i.e., $n_1 = n_2$). We refer this model as the α-unproportional processor sharing model.

Result 3.5 *A product-form solution applies for the α-unproportional processor sharing model of the form* (6)*, with*

$$H(\mathbf{n}) = \left[\prod_{i=1}^{2} \lambda_i^{n_i}\right] P(\mathbf{n}) \left(\frac{\alpha}{1-\alpha}\right)^{\max(n_1,n_2)} \left(\frac{1}{\alpha}\right)^{n_1+n_2}, \text{ for } \mathbf{n} \in C. \tag{37}$$

Proof: To show that this model possesses a product-form solution we need to investigate in verifying condition (8) or equivalently (11) so that Theorem 2.1 applies. For this model it suffices to investigate in the following cycles to verify the condition:

$$\begin{aligned} p &= \mathbf{n} \to \mathbf{n}+\mathbf{e}_1 \to \mathbf{n}+\mathbf{e}_1+\mathbf{e}_2 \to \mathbf{n}+\mathbf{e}_2 \to \mathbf{n}, \\ \bar{p} &= \mathbf{n} \to \mathbf{n}+\mathbf{e}_2 \to \mathbf{n}+\mathbf{e}_1+\mathbf{e}_2 \to \mathbf{n}+\mathbf{e}_1 \to \mathbf{n}. \end{aligned} \tag{38}$$

These cycles need to be considered for the following five scenarios: $n_1 = n_2$, $n_1 + 1 = n_2$, $n_1 - 1 = n_2$, $n_1 + 1 > n_2$ and $n_1 - 1 < n_2$, respectively. For these scenarios the transition rates differ, due to the specific sharing function defined in Equation (36). For the three state space regions where the sharing function differs the products of the transition rates for the paths p and $\bar{p}$, as in Equation (38), equal:

$$\begin{aligned} \theta(p) &= \bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_1)\bar{q}(\mathbf{n}+\mathbf{e}_1, \mathbf{n}+\mathbf{e}_1+\mathbf{e}_2)\bar{q}(\mathbf{n}+\mathbf{e}_1+\mathbf{e}_2, \mathbf{n}+\mathbf{e}_2)\bar{q}(\mathbf{n}+\mathbf{e}_2, \mathbf{n}) \\ &= \alpha^2(1-\alpha)^2\Phi(n_1+n_2+2)\Phi(n_1+n_2+1), \\ \theta(\bar{p}) &= \bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_2)\bar{q}(\mathbf{n}+\mathbf{e}_2, \mathbf{n}+\mathbf{e}_1+\mathbf{e}_2)\bar{q}(\mathbf{n}+\mathbf{e}_1+\mathbf{e}_2, \mathbf{n}+\mathbf{e}_1)\bar{q}(\mathbf{n}+\mathbf{e}_1, \mathbf{n}) \\ &= \alpha^2(1-\alpha)^2\Phi(n_1+n_2+2)\Phi(n_1+n_2+1). \end{aligned}$$

Thus Equation (8) is fulfilled since for all scenarios $\theta(p) = \theta(\bar{p})$. Next note that Equation (37) is obtained following the same lines as in the example given in Section 3.1 or equivalently by Equation (7). The result (37) then follows by substitution of Equations (4) and (36) in Equation (7), with $H(\mathbf{n})$ as defined in Equation (37) and proper scaling. This completes the proof. □

Remark 3.6 *Note that for this example the station with the highest workload receives more capacity than the other station. When the stations have equal workload, both receive an equal share of the total capacity. But since α can be arbitrarily close to 0, not all capacity needs to be used if the workloads are equal. Thus, as a price to pay to satisfy the invariance condition (11) note that a capacity of α is lost when $n_1 = n_2$. It is remarkable that this model possesses a product-form solution, since it is not work-conserving in the state $n_1 = n_2$.*

3.4 Example: State space restriction

In general, state space restrictions of a model that possesses a product-form solution do not necessary possesses a product-form solution itself. However, it is known from [27] that a model, which is reversible itself, possesses a product form at any state space C also possesses a product form at any coordinate convex state space, where coordinate convex is defined by:

$$\mathbf{n} \in C \Rightarrow \mathbf{n} - \mathbf{e}_i \in C, \text{ for } i = 1,\ 2. \tag{39}$$

The proof is stated in Theorem 1 of In [27], namely that the state distribution holds for arbitrary resource sharing policies.

We give an example of a coordinate convex state space restriction for forward reference, since comparing a similar state space restriction for the parallel and the tandem model (see Section 4.4 below) leads to remarkable observations. To this end, consider in this example the service and blocking functions as given in Section 3.1. We restrict the state space of this model by elimination of all states $\mathbf{n}$ with $n_1 \leq n_2$, which can be enforced by $b_1(\mathbf{n}) = 0$ for $n_1 \leq n_2$. This leads to the following coordinate convex state space:

$$C = \{\mathbf{n} \mid n_1 \geq n_2 - 1, n_2 \geq 0\}. \tag{40}$$

This state space restriction is presented in the left figure of Figure 4. We illustrate that the product-form solution indeed holds by verifying Equation (8) for the paths

$$\begin{aligned} p &= (0,2) \to (1,2) \to (1,3) \to (0,3) \to (0,2), \\ \bar{p} &= (0,2) \to (0,3) \to (1,3) \to (1,2) \to (0,2). \end{aligned}$$

And, indeed $\theta(p) = \theta(\bar{p})$ holds, since with (for convenience) $\Phi(n_1 + n_2) = 1$ for all $\mathbf{n} \in C$ and $\lambda_i = 1$, for $i = 1,\ 2$;

$$\begin{aligned} \theta(p) &= 1 \cdot 1 \cdot (3/4) \cdot (1/3) = 1/4, \\ \theta(\bar{p}) &= 1 \cdot 1 \cdot (3/4) \cdot (1/3) = 1/4. \end{aligned}$$

3.5 Example: Two-station limited PS-model

Now we consider the two-station extension of the limited processor sharing (LPS) queue, recently studied in [34,26,42–44], which works as follows: Instead of taking all jobs immediately in service and share the capacity among all these jobs, we consider that $k_i(\mathbf{n})$ jobs receive service and that $k_i(\mathbf{n})$ is bounded by c_i. If there are more than c_i jobs in station i these jobs has to wait until one of the $k_i(\mathbf{n})$ jobs its service is completed. This is defined by the following sharing function:

$$\begin{aligned} s_1(\mathbf{n}) &= k_1(\mathbf{n})/(k_1(\mathbf{n}) + k_2(\mathbf{n})), \quad k_1(\mathbf{n}) = \min(n_1, c_1), \\ s_2(\mathbf{n}) &= k_2(\mathbf{n})/(k_1(\mathbf{n}) + k_2(\mathbf{n})), \quad k_2(\mathbf{n}) = \min(n_2, c_2). \end{aligned} \tag{41}$$

Note that each station receives a fraction of the capacity based on the number of jobs in both stations, and not as in recently studied LPS models shared

among only jobs in one station. Let c_1 and/or c_2 be finite and let the state space be defined as all non-negative integer values for n_1 and n_2 which is:

$$C = \{\mathbf{n} \mid 0 \leq n_i,\ i = 1,\ 2\}, \tag{42}$$

This model is illustrated in Figure 1 where $c_i = 3$ for $i = 1,\ 2$ and where $n_1 = 4$ and $n_2 = 2$. Thus one job in the first station is not in service, but is in the queue, and remains in the queue until one of three jobs in service leaves the station.

Result 3.7 *The two-station limited parallel PS-model violates a product-form solution.*

Proof: The proof is based on a counter-example, so that Equation (13) holds, and equivalent, Equation (8) or Equation (11) does not hold. For this, let $\Phi(k) = 1$ for all $k \geq 1$. Note that the routing is again reversible, and we verify if the products of the transition rates of the cycles satisfy Equation (13) such that the adjoint model is reversible. Consider the limited processor sharing model with $c_1 = 2$ and $c_2 = 3$. Based on verifying Equation (13) we construct the following paths p_1 and p_2:

$$\begin{aligned}
p_1 &= (4,3) \to (4,2) \to (4,1) \to (3,1) \to (2,1) \to (1,1),\\
\bar{p}_1 &= (1,1) \to (2,1) \to (3,1) \to (4,1) \to (4,2) \to (4,3),\\
p_2 &= (4,3) \to (3,3) \to (2,3) \to (1,3) \to (1,2) \to (1,1),\\
\bar{p}_2 &= (1,1) \to (1,2) \to (1,2) \to (2,3) \to (3,3) \to (4,3).
\end{aligned}$$

Take $\lambda_1 = 1$ and $\lambda_2 = 1$. This brings us to the following values of $\Theta(p_i)$ as in (13),

$$\begin{aligned}
\Theta(p_1) &= \theta(p_1)/\theta(\bar{p}_1) = (2/5) \cdot (2/5) \cdot (3/4) \cdot (3/4) \cdot (2/3) = 3/50,\\
\Theta(p_2) &= \theta(p_1)/\theta(\bar{p}_1) = (3/5) \cdot (3/5) \cdot (2/4) \cdot (2/3) \cdot (2/3) = 4/50.
\end{aligned}$$

Thus note that $\Theta(p_1) \neq \Theta(p_2)$. Hence, the necessary reversibility condition (13) is violated, and thus no product form exists. □

Remark 3.8 *Most remarkably, a single-station limited processor sharing queue obviously has a product form, but the structure of the network, in which the sharing depends on the state of the entire model, does not. Because of the limiting number of jobs in service, the order of arrival of the jobs becomes leading, since a job not into service can not be exchanged for a job in service, due to the fact that the service speed does not only depend on that station itself, but also on the other station. This dependency of stations results in the stringent order of the jobs, which results in violation of the reversibility conditions.*

A way to retain the product form for the two-station limited processor sharing parallel model is to restrict the state space artificially such that there can never be more than c_i jobs in station i for $i = 1,\ 2$. The following access blocking functions give a proper state space restriction with respect to the existence of a product-form solution:

$$b_1(\mathbf{n}) = 0 \text{ if } n_1 \geq c_1, \tag{43}$$
$$b_2(\mathbf{n}) = 0 \text{ if } n_2 \geq c_2. \tag{44}$$

These access blocking functions limit the state space to

$$C = \{\mathbf{n} \mid 0 \leq n_i \leq c_i,\ i = 1,\ 2\}. \tag{45}$$

Thus, if a job arrives at a station i while there are already c_i jobs present, then this job is blocked. This model is referred to as the truncated two-station limited parallel PS-model.

Result 3.9 *The truncated two-station limited parallel PS-model possesses a product form of the form* (6) *with* $H(\mathbf{n})$ *as in Equation* (20)*.*

Proof: We again rely on Theorem 2.1 to prove the existence of the product form and its specific form (20). Observe that $s_i(\mathbf{n})$ is equally defined as in the natural processor sharing form (17) for the state space C in Equation (45) and that also on the boundaries the routing remains reversible and transitions are similarly defined as in Section 3.4. This leads immediately to the conclusion that the product form (6) applies, since Equation (20) suffices, which can be verified analogue to the proof in Section 3.4. The form of the product form, (20), follows due to the previous observation, following the lines in Section 3.4. This completed the proof. □

The state space restriction of Section 3.4 in Equation (40) and the state space truncation (45) of the example in this section are illustrated by Figure 4.

Remark 3.10 *Results for showing that a product form cannot hold appear to be rare in the literature. From [7] such results can be deducted if a proper transformation is made, however in the present setting it follows directly. The observation that a model does not have a product form is very important, and can lead to adjustments of the model such that a product form still applies. Note that these adjusted models can be used to develop approximations for the steady-state distribution of non-product form models and can be used to derive error bounds (which falls beyond the scope of the present paper).*

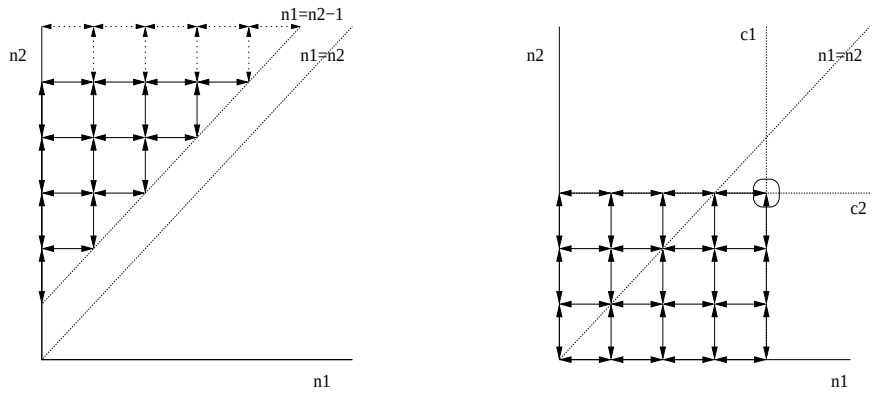

Fig. 4. Parallel Model: The left figure illustrates state space (40) and the right figure illustrates state space (45). For both truncations the product form (6) applies. In the right figure the state (c_1, c_2) is marked for further reference in Section 4.4.

4 Tandem Model

In this section we apply Theorem 2.1 to the tandem model, by using similar examples as for the parallel model. In this model the arrivals at the second station are fed by departures from the first station as described in Section 2.2. We show the similarities and differences for the parallel and tandem models with respect to product-form solutions.

4.1 Example: Proportional PS-model

Consider the following two-station tandem model with the total capacity equals $\Phi(n_1 + n_2)$, and the fraction $s_i(\mathbf{n})$ of this capacity allocated to the stations as given in Equation (17), and thus the capacity is shared proportional the number of jobs in each of the stations. Let the state space C be as in Equation (18). We refer to this model as the proportional tandem PS-model.

Result 4.1 *The proportional tandem PS-model possesses a product-form solution of the form* (6)*, with*

$$H(\mathbf{n}) = \lambda^{n_1+n_2} P(\mathbf{n}) \binom{n_1 + n_2}{n_1}, \text{ for } \mathbf{n} \in C. \tag{46}$$

Proof: To prove this result we show that Theorem 2.1 applies, by verifying Equation (8). Therefore we construct the adjoint Markov chain. The transition

rates of the original Markov chain are as follows:

$$\begin{aligned} q(\mathbf{n}, \mathbf{n}+\mathbf{e}_1) &= \lambda, \\ q(\mathbf{n}, \mathbf{n}-\mathbf{e}_1+\mathbf{e}_2) &= \tfrac{n_1}{n_1+n_2}\Phi(n_1+n_2), \\ q(\mathbf{n}, \mathbf{n}-\mathbf{e}_2) &= \tfrac{n_2}{n_1+n_2}\Phi(n_1+n_2). \end{aligned} \tag{47}$$

Note that the routing of this model is not reversible, and thus we construct the adjoint Markov chain transition rates according to Equation (5), which results in the following rates:

$$\begin{aligned} \bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_1) &= \lambda, \\ \bar{q}(\mathbf{n}, \mathbf{n}-\mathbf{e}_1+\mathbf{e}_2) &= \tfrac{n_1}{n_1+n_2}\Phi(n_1+n_2), \\ \bar{q}(\mathbf{n}, \mathbf{n}-\mathbf{e}_2) &= \tfrac{n_2}{n_1+n_2}\Phi(n_1+n_2), \\ \bar{q}(\mathbf{n}+\mathbf{e}_1, \mathbf{n}) &= \tfrac{n_1}{n_1+n_2}\Phi(n_1+n_2), \\ \bar{q}(\mathbf{n}-\mathbf{e}_1+\mathbf{e}_2, \mathbf{n}) &= \tfrac{n_2}{n_1+n_2}\Phi(n_1+n_2), \\ \bar{q}(\mathbf{n}-\mathbf{e}_2, \mathbf{n}) &= \lambda. \end{aligned} \tag{48}$$

Similar to the parallel model, for the tandem model any cycle can be built from just two basic cycles, and therefore, it suffices to consider only the following two basic cycles:

$$\begin{aligned} p_1 &= \mathbf{n} \to \mathbf{n}+\mathbf{e}_1 \to \mathbf{n}-\mathbf{e}_1+\mathbf{e}_2 \to \mathbf{n}, \\ p_2 &= \mathbf{n} \to \mathbf{n}+\mathbf{e}_2 \to \mathbf{n}+\mathbf{e}_2-\mathbf{e}_1 \to \mathbf{n}. \end{aligned} \tag{49}$$

Substituting the expressions in Equation (48) in Equation (10) we obtain:

$$\begin{aligned} \theta(p_1) &= \bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_1)\bar{q}(\mathbf{n}+\mathbf{e}_1, \mathbf{n}-\mathbf{e}_1+\mathbf{e}_2)\bar{q}(\mathbf{n}-\mathbf{e}_1+\mathbf{e}_2, \mathbf{n}) \\ &= \lambda \cdot \tfrac{(n_1+1)}{n_1+n_2+1}\Phi(n_1+n_2+1) \cdot \tfrac{(n_2+1)}{n_1+n_2}\Phi(n_1+n_2+1), \\ \theta(\bar{p}_1) &= \bar{q}(\mathbf{n}, \mathbf{n}-\mathbf{e}_1+\mathbf{e}_2)\bar{q}(\mathbf{n}-\mathbf{e}_1+\mathbf{e}_2, \mathbf{n}+\mathbf{e}_1)\bar{q}(\mathbf{n}+\mathbf{e}_1, \mathbf{n}) \\ &= \lambda \cdot \tfrac{(n_2+1)}{n_1+n_2}\Phi(n_1+n_2+1) \cdot \tfrac{(n_1+1)}{n_1+n_2+1}\Phi(n_1+n_2+1), \end{aligned}$$

and verifying Equation (8), indeed $\theta(p_1)$ equals $\theta(\bar{p}_1)$. Using similar arguments we show that $\theta(p_2) = \theta(\bar{p}_2)$. Thus, $\theta(p_i) = \theta(\bar{p}_i)$ for $i = 1,\ 2$, and since from these two cycles any other cycles can be constructed, the reversibility applies (i.e. Equation (8)) and there exists a product-form solution (6). Next, we prove that $H(\mathbf{n})$ has the form (46). This can be obtained by solving Equation (7) recursively starting with $H(\mathbf{0}) = \Phi(0)$, which leads to Equation (27) and Equation (28), where λ_2 is replaced by λ. Next, since the solution of the

recursive scheme also satisfies the third recursive relation

$$H(\mathbf{n}+\mathbf{e}_1)\tfrac{(n_1+1)}{n_1+n_2+1}\Phi(n_1+n_2+1) = H(\mathbf{n}+\mathbf{e}_2)\tfrac{(n_2+1)}{n_1+n_2+1}\Phi(n_1+n_2+1),$$

the local balance equations are satisfied. This completes the proof. □

Remark 4.2 *This result may be surprising since the model can be seen as a single processor sharing model with 2 classes of jobs where after completion of service of a class 1 job it becomes a class 2 job, see in [7]. However, it illustrates how to apply the theorem with the use of the adjoint Markov chain which is different from the original Markov chain due to the non-reversible routing of the original chain.*

Remark 4.3 *Note that the function $H(\mathbf{n})$ differs only in the routing part from the form given in Equation (20), due to the fact that λ feds both, station 1, and station 2 after completion of service at station 1. Contrary, for the parallel model station 2 is fed by its own arrival process with rate λ_2. The service part in $H(\mathbf{n})$, based on $s_i(\mathbf{n})$, is similar for the parallel and tandem model which was expected since $s_i(\mathbf{n})$ is equally defined, for i=1, 2.*

Remark 4.4 *(Expression for the total population) In the particular proportional case we can also obtain a simple standard geometric-type expression for the steady-state distribution $\pi(\nu)$ for the total number of jobs $\nu = n_1 + n_2$. To this end, by (6) and (46) we obtain:*

$$\begin{aligned}\pi(\nu) &= c \sum_{n_1,n_2:n_1+n_2=\nu} \lambda^{n_1+n_2}\binom{n_1+n_2}{n_1}\left[\prod_{k=1}^{\nu}\Phi(k)\right]^{-1}\left(\frac{1}{\mu_1}\right)^{n_1}\left(\frac{1}{\mu_2}\right)^{n_2}\\ &= c\lambda^{\nu}\left[\prod_{k=1}^{\nu}\Phi(k)\right]^{-1}\left(\frac{1}{\mu_1}+\frac{1}{\mu_2}\right)^{\nu}\\ &= c(\lambda\tau)^{\nu}\left[\prod_{k=1}^{\nu}\Phi(k)\right]^{-1},\end{aligned}$$

with $\tau = \left(\frac{1}{\mu_1}+\frac{1}{\mu_2}\right)$ the total mean service time.

For the parallel model from Section 3.1, we similarly find by (6) and (20):

$$\pi(\nu) = c(\lambda\tau)^{\nu}\left[\prod_{k=1}^{\nu}\Phi(k)\right]^{-1}, \tag{50}$$

with

$$\tau = \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \frac{1}{\mu_1} + \frac{\lambda_2}{\lambda_1 + \lambda_2} \frac{1}{\mu_2} \right),$$
$$\lambda = \lambda_1 + \lambda_2.$$

Hence, for both the parallel and the tandem model under proportional sharing and c a normalizing constant we obtain:

$$\pi(\nu) = c\rho^{\nu} \left[\prod_{k=1}^{\nu} \Phi(k) \right]^{-1}, \tag{51}$$

with ρ the mean workload with $\rho = \lambda(\beta_1 + \beta_2)$ for the tandem model, and $\rho = \lambda_1\beta_1 + \lambda_2\beta_2$ for the parallel model and with $\beta_i = \frac{1}{\mu_i}$.

Remark 4.5 *Note that Equation (51) can be in fact be seen as a simple insensitivity result, i.e. that the expression does not depends on the routing mechanism but only on the average workload ρ. As such this insensitivity result is in line (though somewhat different) with more standard insensitivity results from processor sharing systems, e.g. [5,7]. Note, however, that the (insensitivity expression (51) also applies without assuming a strict processor sharing discipline, i.e. in which the service at a station is equally spread over all jobs. We know that such an insensitivity result for examples as in Sections 4.2 to 4.6 does not seem to can be concluded since expression (51) essentially uses the multinomial coefficient.*

4.2 Example: Unproportional PS-model with full capacity to one station

As for the parallel model, for the tandem model we also continue with unproportional processor sharing examples. First we consider the example where the full capacity is always allocated to one station by setting the capacity of the other station at value 0, and in the next section we consider the α-unproportional model. For the model considered in this section, the model where the full capacity is assigned to the station with the highest workload, recall the sharing function $s_i(\mathbf{n})$ as in Equation (30) and let the blocking function $b_1(\mathbf{n})$ be equal to the blocking function for the station 1 as given in Equation (29). This model is referred as the unproportional tandem PS-model. Because of sharing and blocking functions the state space C is defined as in Equation (31). Note that blocking can not occur at the second station, however, we obtain the same state space as for the tandem model. In the left figure of Figure 5 this state space is illustrated.

Result 4.6 *The unproportional tandem PS-model possesses a product-form*

solution of the form (6)*, with*

$$H(\mathbf{n}) = \lambda^{n_1+n_2} P(\mathbf{n}), \text{ for } \mathbf{n} \in \mathbf{C}. \tag{52}$$

Proof: We construct the adjoint Markov chain according to Equation (5) so that we can verify Equation (8) and rely on Theorem 2.1. The transition rates of the adjoint Markov chain are as follows:

$$\begin{aligned}
\bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_1) &= \lambda, \\
\bar{q}(\mathbf{n}+\mathbf{e}_1, \mathbf{n}+\mathbf{e}_2) &= \Phi(n_1+n_2), \\
\bar{q}(\mathbf{n}+\mathbf{e}_2, \mathbf{n}) &= \Phi(n_1+n_2), \\
\bar{q}(\mathbf{n}+\mathbf{e}_1, \mathbf{n}) &= \Phi(n_1+n_2), \\
\bar{q}(\mathbf{n}+\mathbf{e}_2, \mathbf{n}+\mathbf{e}_1) &= \Phi(n_1+n_2), \\
\bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_2) &= \lambda,
\end{aligned} \tag{53}$$

where the transition rates only exist for given C as stated in Equation (31). Along the lines of the previous proof, we again need to verify for only two cycles if Equation (8) is satisfied, because the other cycles can be constructed with these cycles. Consider the cycles

$$\begin{aligned}
p &= \mathbf{n} \to \mathbf{n}+\mathbf{e}_1 \to \mathbf{n}+\mathbf{e}_2 \to \mathbf{n}, \\
\bar{p} &= \mathbf{n} \to \mathbf{n}+\mathbf{e}_2 \to \mathbf{n}+\mathbf{e}_1 \to \mathbf{n}.
\end{aligned}$$

For these cycles we use Equation (53) in Equation (10) which results in the following:

$$\begin{aligned}
\theta(p) &= \bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_1)\bar{q}(\mathbf{n}+\mathbf{e}_1, \mathbf{n}+\mathbf{e}_2)\bar{q}(\mathbf{n}+\mathbf{e}_2, \mathbf{n}), \\
&= \lambda \cdot \Phi(n_1+n_2+1) \cdot \Phi(n_1+n_2+1), \\
\theta(\bar{p}) &= \bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_2)\bar{q}(\mathbf{n}+\mathbf{e}_2, \mathbf{n}+\mathbf{e}_1)\bar{q}(\mathbf{n}+\mathbf{e}_1, \mathbf{n}), \\
&= \lambda \cdot \Phi(n_1+n_2+1) \cdot \Phi(n_1+n_2+1).
\end{aligned}$$

And indeed $\theta(p) = \theta(\bar{p})$ for these two cycles and thus for all cycles p in C. Notice that Equation (52) can be obtained recursively solving Equation (7), starting with $H(\mathbf{0}) = \Phi(0)$, which results in Equation (34) with λ_2 replaced by λ. This recursion leads to the form given in Equation (52). This completes the proof. □

The left figure in Figure 5 illustrates the state space restrictions as in Equation (31), however, also other state-space restrictions for which (52) applies

are possible, an example is illustrated in the right figure. Note the triangular structure of the cycles.

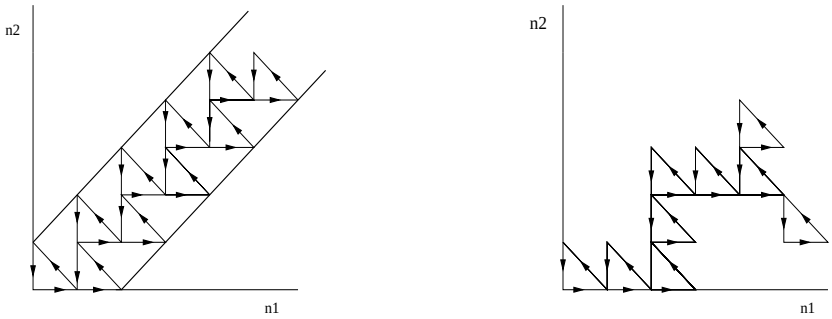

Fig. 5. Tandem Model: Figures for state space C for which the product form (6) applies with positive transition rates indicated by an arrow (all other rates are equal to 0).

4.3 Example: α-Unproportional PS-model

Now, we consider again unproportional and non-zero sharing functions $s_i(\mathbf{n})$ over both stations, as in Equation (36) on the state space C (18). This model is called the α-unproportional tandem PS-model. This model is presented in [12], and for a special value of α some results are presented. However, in this Section we explain explicitly how to obtain the product form and compare this with the parallel version of the model. To this end, we observe that the α-unproportional tandem model still retains the necessary invariance (11) or equivalently (8) to secure a product form, for example with arbitrary $0 < \alpha < 1/2$. This model is illustrated in Figure 6, wherein only a few cycles are presented, representing the different rates depending on the state (n_1, n_2). Since there are no limitations on the state space, the complete state space is filled with these triangular structured transition rates.

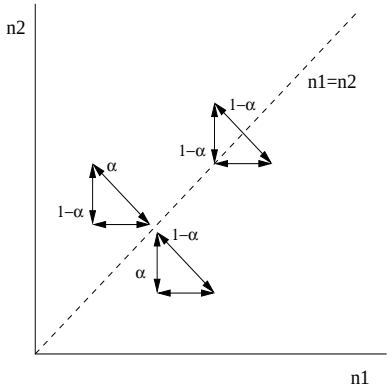

Fig. 6. Tandem Model: The α-unproportional PS-model. In this figure in each of the regions ($n_1 < n_2$, $n_1 = n_2$, and $n_1 > n_2$) the transition rates of a cycle are given.

Result 4.7 *The α-unproportional tandem PS-model possesses a product-form solution with the fraction of capacity allocated according to Equation* (36) *of the form* (6)*, with*

$$H(\mathbf{n}) = \lambda^{n_1+n_2} P(\mathbf{n}) \left(\frac{\alpha}{1-\alpha}\right)^{\max(n_1,n_2)} \left(\frac{1}{\alpha}\right)^{n_1+n_2}, \text{ for } \mathbf{n} \in C. \tag{54}$$

Proof: For the proof we follow similar lines as the proof given in Section 3.3. To this end, we construct the adjoint Markov chain according to Equation (5) to verify condition (8). The adjoint transition rates become:

$$\begin{aligned}
\bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_1) &= \lambda, \\
\bar{q}(\mathbf{n}+\mathbf{e}_1, \mathbf{n}) &= (1-\alpha), \quad \text{if } n_1+1 < n_2, \\
\bar{q}(\mathbf{n}+\mathbf{e}_1, \mathbf{n}) &= \alpha, \quad \text{if } n_1+1 \geq n_2, \\
\bar{q}(\mathbf{n}+\mathbf{e}_2, \mathbf{n}) &= \alpha, \quad \text{if } n_2+1 \leq n_1, \\
\bar{q}(\mathbf{n}+\mathbf{e}_2, \mathbf{n}) &= (1-\alpha), \quad \text{if } n_2+1 > n_1, \\
\bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_2) &= \lambda, \\
\bar{q}(\mathbf{n}+\mathbf{e}_1, \mathbf{n}+\mathbf{e}_2) &= (1-\alpha), \quad \text{if } n_1+1 > n_2, \\
\bar{q}(\mathbf{n}+\mathbf{e}_1, \mathbf{n}+\mathbf{e}_2) &= \alpha, \quad \text{if } n_1+_1 \leq n_2, \\
\bar{q}(\mathbf{n}+\mathbf{e}_2, \mathbf{n}+\mathbf{e}_1) &= \alpha, \quad \text{if } n_2+1 \leq n_1, \\
\bar{q}(\mathbf{n}+\mathbf{e}_2, \mathbf{n}+\mathbf{e}_1) &= (1-\alpha), \quad \text{if } n_2+1 > n_1.
\end{aligned} \tag{55}$$

The adjoint transition rates differ, depending on n_1 and n_2. We only verify Equation (8) for the three cycles illustrated in Figure 6, noting that using these cycles, all cycles in the state space C can be constructed. Let

$$\begin{aligned}
p &= \mathbf{n} \to \mathbf{n}+\mathbf{e}_1 \to \mathbf{n}+\mathbf{e}_2 \to \mathbf{n}, \\
\bar{p} &= \mathbf{n} \to \mathbf{n}+\mathbf{e}_2 \to \mathbf{n}+\mathbf{e}_1 \to \mathbf{n}.
\end{aligned}$$

The product of the transition rates for these paths are as follows; for $n_1 > n_2$,

$$\begin{aligned}
\theta(p) &= \lambda \cdot \alpha\Phi(n_1+n_2+1) \cdot (1-\alpha)\Phi(n_1+n_2+1), \\
\theta(\bar{p}) &= \lambda \cdot (1-\alpha)\Phi(n_1+n_2+1) \cdot \alpha\Phi(n_1+n_2+1),
\end{aligned}$$

and for $n_1 < n_2$,

$$\begin{aligned}
\theta(p) &= \lambda \cdot (1-\alpha)\Phi(n_1+n_2+1) \cdot \alpha\Phi(n_1+n_2+1), \\
\theta(\bar{p}) &= \lambda \cdot \alpha\Phi(n_1+n_2+1) \cdot (1-\alpha)\Phi(n_1+n_2+1),
\end{aligned}$$

and for $n_1 = n_2$,

$$\theta(p) = \lambda \cdot (1-\alpha)\Phi(n_1+n_2+1) \cdot (1-\alpha)\Phi(n_1+n_2+1),$$
$$\theta(\bar{p}) = \lambda \cdot (1-\alpha)\Phi(n_1+n_2+1) \cdot (1-\alpha)\Phi(n_1+n_2+1).$$

And thus indeed $\theta(p) = \theta(\bar{p})$ for all paths in C. This leads to the conclusion that for this model the product-form solution exists and is given with $H(\mathbf{n})$ as in Equation (37). Since the state space equals the state space of the parallel model, and the model has the same sharing function, we immediately conclude that $H(\mathbf{n})$ has the same form as the parallel model up to the routing part, as already can be seen in the proof given in Section 4.1. Thus the proof is completed. □

Remark 4.8 *Note that for this model less cycles need to be checked as for the parallel model, due to the routing structure. But the results are equivalent, both examples lead to a product-form solution.*

4.4 Example: State space restriction

As we have seen in previous sections for the tandem model, the tandem model and the parallel have many similarities with respect to the obtained product form for the same sharing functions and blocking functions. In the following examples of the tandem model we observe that differences occur. Consider the sharing function as in Equation (17) and state space C (40). For the tandem model, with proportional sharing of the capacity, the sharing function needs to be supplemented with:

$$s_2(\mathbf{n}) = 0, \text{ for } n_1 = n_2 - 2, \tag{56}$$

and similarly the following additionally to the blocking functions in Equation (40) is needed:

$$b_1(\mathbf{n}) = 1, \text{ for } n_1 \geq n_2. \tag{57}$$

In the left figure of Figure 8 the transitions for this particular proportional sharing model in the state space C defined in (40) are illustrated.

Result 4.9 *The standard PS model has a product form*

$$\pi(\mathbf{n}) = c\lambda^{n_1+n_2} P(\mathbf{n}) \binom{n_1+n_2}{n_1}, \tag{58}$$

for state space C as given in (40)*, supplemented with the sharing function in Equation* (56) *and the blocking function in Equation* (57)*.*

Proof: For the proof of this result we refer to Section 4.1. We only have to verify if Equation (8) is satisfied for the cycles in the admissible state space C (40). Note that we block service at the second station for the states $n_1 = n_2 - 2$ to obtain the local balance for the second station (which can be verified by substitution of the transition rates (48) of the adjoint Markov chain in the Kolmogorov equations (3)). We now conclude similarly to the example in Section 4.1 that the product-form solution applies with the same $H(\mathbf{n})$, wherein $H(\mathbf{n})$ can be obtained following the lines of the proof of Result 4.1). □

Remark 4.10 *Note that these results cannot be concluded from product-form results by simply restricting the state space under reversibility conditions such as in [35], since transition rates in the coordinate convex state space need to be changed such that reversibility of the adjoint Markov chain remains.*

Remark 4.11 *Most remarkable is the following: to obtain a product form for the tandem model similarly to the parallel model different sharing functions are necessary such that the station balance equations are verified. For the tandem model additionally the service from the second station needs to be blocked in the state space, for the case $n_1 = n_2 - 2$, which results in function $H(\mathbf{n})$ similar up to the routing part, the part containing the λ's.*

4.5 Example: Two-station limited PS-model

Now, we consider the limited tandem PS-model, with sharing function $s_i(\mathbf{n})$ as in Equation (41). This function shares the capacity similar to the model in the previous section in the inner region of the state space C, and differs as soon as the boundaries are reached, namely if $n_1 = c_1$ or $n_2 = c_2$. We consider the state space

$$C = \{\mathbf{n} \mid n_1, n_2 \geq 0\}. \tag{59}$$

Result 4.12 *A product form is necessarily violated for the two-station limited tandem PS-model.*

Proof: This can be proved by verification of Equation (8). To this end, let $\Phi(n_1 + n_2) = 1$ and let $\lambda = 1$. The transitions of the adjoint Markov chain are shown in Figure 7 and, more precisely, in formula as follows:

$$\begin{aligned}
\bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_1) &= \bar{q}(\mathbf{n}, \mathbf{n}+\mathbf{e}_2) = 1, \\
\bar{q}(\mathbf{n}+\mathbf{e}_1, \mathbf{n}+\mathbf{e}_2) &= \bar{q}(\mathbf{n}+\mathbf{e}_1, \mathbf{n}) = s_1(\mathbf{n}+\mathbf{e}_1), \\
\bar{q}(\mathbf{n}+\mathbf{e}_2, \mathbf{n}+\mathbf{e}_1) &= \bar{q}(\mathbf{n}+\mathbf{e}_2, \mathbf{n}) = s_2(\mathbf{n}+\mathbf{e}_2).
\end{aligned} \tag{60}$$

Using a counter-example, we show that Equation (8) is necessarily violated. Consider $c_1 = 2$ and $c_2 = \infty$ as also presented in Figure 7 and consider the following cycles:

$$\begin{aligned}
p &= (1,1) \to (2,1) \to (3,1) \to (3,2) \to (2,2) \to (1,2) \to (1,1), \\
\bar{p} &= (1,1) \to (1,2) \to (2,2) \to (3,2) \to (3,1) \to (2,1) \to (1,1),
\end{aligned}$$

and the products of these paths according to Equation (10) equal

$$\begin{aligned}
\theta(p) &= 1 \cdot 1 \cdot 1 \cdot (1/2) \cdot (1/2) \cdot (2/3) = 1/6, \\
\theta(\bar{p}) &= 1 \cdot 1 \cdot 1 \cdot (1/2) \cdot (2/3) \cdot (2/3) = 2/9,
\end{aligned}$$

and indeed $\theta(p) \neq \theta(\bar{p})$. As a consequence, condition (8) and thus also the necessary reversibility condition is violated so that the product form (6) fails. Similar (counter) examples can be given for any finite c_1 and/or c_2. □

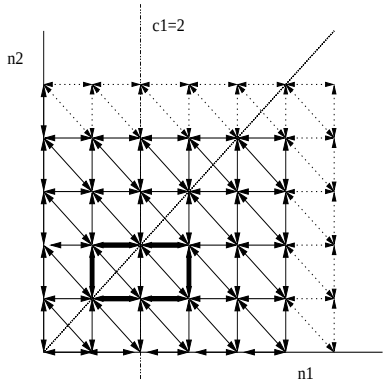

Fig. 7. Tandem Model: The adjoint transition rates for the limited processor sharing model.

4.6 Example: Truncated two-station limited PS model

In line with Section 4.4 and as an extension to station interdependent processor sharing services of a product-form modification result in [10] for independent services, a way to retain the product form for the case where $k_i(\mathbf{n})$ (the number of jobs in service at station i) is limited, is to restrict the state space such that there can be no more than c_1 jobs at station 1 and c_2 at station 2. This

idea is already introduced for the parallel model.

However, to secure the necessary reversibility for the adjoint Markov chain additional boundary conditions are required, namely,

$$\begin{aligned} &s_1(\mathbf{n}) = 0, \text{ if } n_2 = c_2, \\ &s_2(\mathbf{n}) = 0, \text{ if } n_1 = c_1, \\ &b_1(\mathbf{n}) = 0, \text{ if } n_1 = c_1 \text{ or } n_2 = c_2. \end{aligned} \tag{61}$$

These additional boundary conditions limit the state space to

$$C = \{\mathbf{n} \mid 0 \leq n_i \leq c_i \quad i = 1,\ 2\}. \tag{62}$$

Result 4.13 *The truncated two-station limited tandem PS-model possesses a product form* (6)*, with* $H(\mathbf{n})$ *as in* (46)*.*

Proof: To prove that the model possesses a product form we rely on the proof in Section 4.1. The model, as defined above, implies reversible routing of the adjoint Markov chain, which can be easily verified by checking the station balance equations (3). Next verifying Equation (8) leads to the same products as in Section 4.1, and similarly to Section 4.1 we obtain $H(\mathbf{n})$. Thus we conclude immediately that the product form exists with $H(\mathbf{n})$ as by (46). □

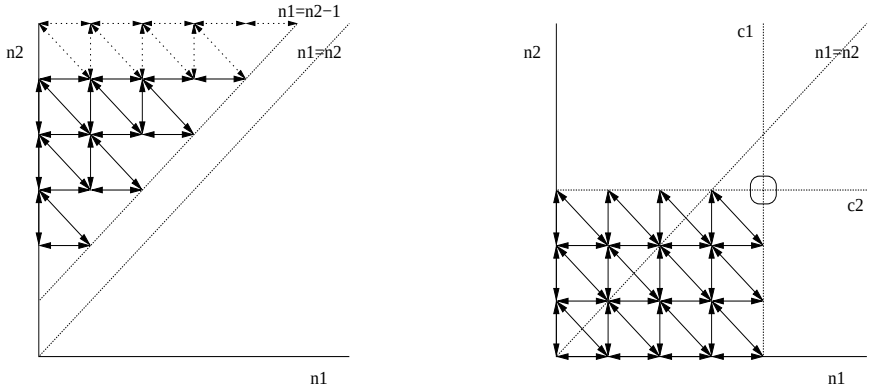

Fig. 8. Tandem Model: The left figure illustrates state space (40), and the right figure illustrates state space (62). For both truncations a product form (6) applies.

Remark 4.14 *In this example we adjusted the sharing function to satisfy the station balance equations. This adjustment leads to a smaller state space then the state space of the parallel model, since the upper-right corner* (c_1, c_2) *can not be reached, because this upper-right corner will ruin the reversibility of the adjoint Markov chain. Thus, the product form has the same form as the parallel model up to the routing part, but the state space differs to let the model possess a product-form solution.*

5 Conclusion and further research

In this paper we extended the product-form results of [10] to a general setting wherein blocking is allowed, as well as state dependent service and fully stopped service. We present an approach for showing whether a model possesses a product form or not, unifying the tandem and parallel model. Illustrative and new examples are presented, wherein remarkable results are presented.

The results presented in Sections 3 and 4 lead to a number of remarkable observations, which will be discussed in more detail below.

First we observe that the product-form results for the parallel and the tandem model have the same form for some examples up to the routing part (λ, versus λ_1 and λ_2). The routing part can be easily explained due to the fact that the input to the second station is fed by λ_2 in the parallel model, and by λ in the tandem model. For equivalent sharing functions and state spaces C, with C as defined in Equation (18) this is the case. However, the product forms do not have the same structure if the state space is truncated, due to blocking of arrivals or stopping of service, such as in the examples in Sections 3.4, 4.4, and 3.6, 4.6. For some examples similar functions of $H(\mathbf{n})$ are obtained, due to proper choice of the state space and stopping some transitions such that the reversibility, and the station balance equations are verified. Thus, although the models are fundamentally different the function $H(\mathbf{n})$ is the same (except the routing part) for both models.

Second, it is remarkable that the two station version of the limited PS-model does not lead to a product form, for the tandem as well as for the parallel model. Since if the service discipline is independent of the station a simple product-form solution applies and even when the capacity is evenly shared among all jobs a product form will suffice. However, if the number of jobs that simultaneously receive service is bounded the structure is ruined. We suspect that this is due to the effect of queueing, which not only depends on the station where the job is served, but also on the other station.

Third, to verify the reversibility condition, we rely on the adjoint Markov chain. It is interesting to observe that this is necessary for the tandem model, but not for the parallel model. To this end, note that the transitions for the tandem model are not reversible, whereas the transition rates of the parallel model are reversible. This illustrates the additional value of defining the adjoint Markov chain, since many model instances fit in the framework presented in Theorem 2.1 to prove that a model does or does not possesses a product-form solution.

The results lead to a number of directions for further research. First, in this paper we focused on networks with two stations, which led to the analysis of two-dimensional state spaces. An interesting area for further research is to investigate to what extent the results can be generalized to models with an arbitrary number of stations. We suspect that this type of generalizations is possible under assumptions about the symmetry of the capacity assignment function $s_i(\mathbf{n})$. For non-symmetrical capacity assignment functions (see for example (17)), additional assumptions are likely to be needed to obtain product-form results. Derivation of this type of generalizations is a challenging area for further research.

Second, the results form an excellent basis for the development of simple yet accurate approximations for the mean sojourn times in case there is no product form. In this context, we may also derive error bounds for these approximations, based on the value-function techniques for related product-form networks [11].

Third, the results provide possibilities for optimization, both for models with and without product-form solutions. We may be able to derive monotonicity and convexity properties of mean sojourn times with respect to the limitations on the number of jobs in service, and with respect to the limitations on the state space. In this context, encouraging monotonicity results have been obtained for the single-station case in [34]. Extensions of these results to the more general setting of the present paper is an interesting area for follow-up research. Another area of interest is the development of efficient strategies for the dynamic assignment of capacities to the stations. To this end, we may study the performance of a control scheme in a Markov decision framework, and consider multi-modularity properties of the value functions. Initial results presented in [40] show that significant performance gains can be obtained by these dynamic schemes compared to state-independent schemes.

References

[1] B. Avi-Itzhak and S. Halfin. Expected response times in a non-symmetric time sharing queue with a limited number of service positions. In *Proceedings of ITC 12*, pages 5.4B.2.1–7, 1988.

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

[3] T. Bonald and L. Massoulié. Impact of fairness on Internet performance. In *Proceedings of ACM Sigmetrics / Performance*, pages 193–209, 2001.

[4] T. Bonald, L. Massoulié, A. Proutière, and J. Virtamo. A queueing analysis of max-min fairness, proportional fairness and balanced fairness. *Queueing Systems*, 53(1-2):65–84, 2006.

[5] T. Bonald and A. Proutière. Insensitivity in processor-sharing networks. *Performance Evaluation*, 49:193–209, 2002.

[6] S.C. Borst, M. Jonckheere, and L. Leskelä. Stability of parallel queueing systems with coupled service rates. *Discrete Events and Stochastic Systems*, 18:447–472, 2008.

[7] K.M. Chandy and A.J. Martin. A characterization of product-form queueing networks. *Journal of the ACM*, 30(2):286–299, 1983.

[8] T. Coenen, H. van den Berg, and R.J. Boucherie. A flow level model for wireless multihop ad hoc networks throughput. In *Proceedings 3rd International Working Conference on Performance Modelling and Evaluation of Heterogeneous Networks (HETNETS)*, volume P34, Ilkley, England, 2005.

[9] J.W. Cohen and O.J. Boxma. *Boundary Value Problems in Queueing System Analysis*. North-Holland, Amsterdam, 1983.

[10] N.M. van Dijk. *Queueing Networks and Product Forms: A Systems Approach*. John Wiley & Sons Ltd., Chichester, England, 1993.

[11] N.M. van Dijk. Bounds and error bounds for queueing networks. *Annals of Operations Research*, 79:295–319, 1998.

[12] N.M. van Dijk. On product form tandem structures. *Mathematical Methods of Operations Research*, 62(3):429–436, 2005.

[13] J. Dilley, R. Friedrich, T. Jin, and J. Rolia. Web server performance measurement and modeling techniques. *Performance Evaluation*, 33(1):5–26, 1998.

[14] G. Fayolle and R. Iasnogorodski. Two coupled processors: The reduction to a Riemann-Hilbert problem. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 47(3):325–351, 1979.

[15] G. Franks, D.C. Petriu, C.M. Woodside, J. Xu, and P. Tregunno. Layered bottlenecks and their mitigation. In *Proceedings of 3rd International Conference on Quantitative Evaluation of Systems*, pages 103–114, 2006.

[16] R. Hariharan, W.K. Ehrlich, P.K. Reeser, and R.D. van der Mei. Performance of Web servers in a distributed computing environment. In *Teletraffic Engineering in the Internet Era*, pages 137–148, 2001. Also Proceedings 17th International Teletraffic Congress (Salvador, December 2001).

[17] M. Harkema, B.M.M. Gijsen, R.D. van der Mei, and Y. Hoekstra. Middleware performance modelling. In *Proceedings international Symposium on Performance Evaluation of Computer and Telecommunication Systems, SPECTS*, pages 733–742, San Jose, CA, 2004.

[18] P. G. Harrison. Turning back time in markovian process algebra. *Theoretical Computer Science*, 290:1947–1986, 2003.

[19] P.G. Harrison. Compositional reversed markov processes with applications to g-networks. *Performance Evaluation*, 57:379–408, 2004.

[20] P.G. Harrison. Reversed processes, product forms and a non-product form. *Journal of Linear Algebra and Applications*, 386:369–381, 2004.

[21] P.G. Harrison and T.T. Lee. Separable equilibrium state probabilities via time-reversal in markovian process algebra. *Theoretical Computer Science*, 346:161–182, 2005.

[22] A. Hordijk and N.M. van Dijk. Networks of queues with blocking. *Performance*, 81:51–65, 1981.

[23] A. Hordijk and N.M. van Dijk. Networks of queues, Part I: Job-local-balance and the adjoint process. Part II: General routing and service characteristics. In *Lecture Notes in Control and Information Sciences*, volume 60, pages 151–205. Springer-Verlag, New York Inc., 1983.

[24] J.R. Jackson. Networks of waiting lines. *Operations Research*, 5:518–521, 1957.

[25] J.R. Jackson. Jobshop-like queueing systems. *Management Science*, 10:131–142, 1963.

[26] M. Jonckheere, R.D. van der Mei, and W. van der Weij. Rate stability and output rates in queueing networks with shared resources. Submitted for publication, 2007.

[27] J.S. Kaufman. Blocking in a shared resource environment. *IEEE Transactions on Communications*, 29(10):1474–1481, 1981.

[28] F.P. Kelly. *Reversibility and Stochastic Networks*. John Wiley & Sons Ltd., New York, 1979.

[29] A. Konheim, I. Meilijson, and A. Melkman. Processor-sharing of two parallel lines. *Journal of Applied Probability*, 18:952–956, 1981.

[30] L. Massoulié and J.W. Roberts. Bandwidth sharing: Objectives and algorithms. In *Proceedings of IEEE INFOCOM Conference*, pages 1395–1403, 1999.

[31] R.D. van der Mei, W.K. Ehrlich, P.K. Reeser, and J.P. Francisco. A decision support system for tuning web servers in distributed object-oriented network architectures. *ACM Performance Evaluation Review*, 27:57–62, 2000.

[32] R.D. van der Mei, R. Hariharan, and P.K. Reeser. Web server performance modeling. *Telecommunication Systems*, 16:361–378, 2001.

[33] M. Miyazawa and P.G. Taylor. A geometric product-form distribution for a queueing network with standard batch arrivals and batch transfers. *Advances in Applied Probability*, 29:523–544, 1997.

[34] M. Nuyens and W. van der Weij. Monotonicity in the limited processor sharing queue. Submitted for publication, 2008.

[35] B. Pittel. Closed exponential networks of queues with saturation. The Jackson-type stationary distribution and its asymptotic analysis. *Mathematics of Operations Research*, 4:357–378, 1979.

[36] S. Ramesh and H.G. Perros. A multilayer client-server queueing network model with synchronous and asynchronous messages. *IEEE Transactions on Software Engineering*, 26(11):1086–1100, 2000.

[37] J.A. Rolia and K.C. Sevcik. The method of layers. *IEEE Transactions on Software Engineering*, 21:689–699, 1995.

[38] R. Schassberger. Insensitivity of steady-state distributions of generalized semi-Markov processes. Part I. *Annals of Applied Probability*, 5:87–99, 1977.

[39] W. van der Weij. Sojourn times in a two-layered tandem queue with limited service positions and a shared processor. Master Thesis, University of Amsterdam, 2004.

[40] W. van der Weij, S. Bhulai, and R.D. van der Mei. Dynamic thread assignment in web server performance optimization. To appear in *Performance Evaluation*, 2008.

[41] C.M. Woodside, J.E. Neilson, D.C. Petriu, and S. Majumdar. The stochastic Rendezvous network model for the performance of synchronous client-server like distributed software. *IEEE Transactions on Computers*, 44:20–34, 1995.

[42] J. Zhang, J.G. Dai, and B. Zwart. Diffusion limits of limited processor sharing queues. Submitted for publication, 2008.

[43] J. Zhang, J.G. Dai, and B. Zwart. Law of large number limits of limited processor sharing queues. Submitted for publication, 2008.

[44] J. Zhang and B. Zwart. Steady state approximations of limited processor sharing queues in heavy traffic. *Queueing Systems*, 60:227–246, 2008.