

Serie Research Memoranda

A generalization of Norton's theorem

R.J. Boucherie

Research Memorandum 1991-61
September 1991



A generalization of Norton's theorem

Richard J. Boucherie and Nico M. van Dijk *

Abstract

A general framework for aggregation and decomposition of product form queueing networks with state dependent routing and servicing is presented. By analogy with electrical circuit theory, the stations are grouped into clusters or subnetworks such that the process decomposes into a global process and a local process. Moreover, the local process factorizes into the subnetworks. The global process and the local processes can be analyzed separately as if they were independent. The global process describes the behaviour of the queueing network in which each cluster is aggregated into a single station, whereas the local process describes the behaviour of the subnetworks as if they are not part of the queueing network. The decomposition and aggregation method formalized in this paper allows us to first analyze the global behaviour of the queueing network and subsequently analyze the local behaviour of the subnetworks of interest or to aggregate clusters into single stations without affecting the behaviour of the rest of the queueing network. Conditions are provided such that:

- the global equilibrium distribution for aggregated clusters has a product form;
- this form can be obtained by merely monitoring the global behaviour;
- the computation of a detailed distribution, including its normalizing constant, can be decomposed into the computation of a global and a local distribution;
- the marginal distribution for the number of jobs at the stations of a cluster can be obtained by merely solving local behaviour.

As a special application, Norton's theorem for queueing networks is extended to queueing networks with state dependent routing such as due to capacity constraints at stations or at clusters of stations and state dependent servicing such as due to service delays for clusters of stations.

Keywords: Aggregation, Blocking, Decomposition, Global traffic equations, Local traffic equations, Norton's theorem, Product form, Queueing networks.

*Address for both authors: Free University, Department of Econometrics, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

1 Introduction

Queueing networks are widely used in computer performance evaluation, telecommunications and manufacturing. Frequently one is interested merely in global characteristics such as the total throughput of a group of stations, sojourn times at parts of the queueing network, transition flows from one part to another or just a global occupancy distribution for clusters of stations rather than a detailed distribution for all stations. Such characteristics can in principle be obtained from the equilibrium distribution of the queueing network. However, the size of the queueing network often prohibits efficient calculation of this equilibrium distribution even if a closed product form expression is available. In such situations one would preferably be able to compute these global characteristics by evaluating or measuring global behaviour. Intuitively this would be achieved by aggregating parts of the queueing network as single queues so that global characteristics can be calculated without calculating the local characteristics. Furthermore, if one is interested in local characteristics of a part of the queueing network only, computation of these characteristics would be simplified if the rest of the queueing network could be aggregated as a single queue or as a small number of single queues. Conversely, as conditions for closed product form expressions can be complicated to verify when state dependent routing and servicing are involved, it would also be appealing when these conditions can be decomposed into global and local conditions. Particularly, detailed product form expressions can then be concluded by merely verifying global and local traffic conditions. Unfortunately, such aggregation and decomposition results are not generally justified as local and global network behaviour is usually integrated by state dependent routing and servicing mechanisms. In [5] an efficient aggregation method similar to Norton's theorem from electrical circuit theory is introduced. Norton's theorem states that

under certain conditions on the structure of the queueing network it is possible to replace a subset of the queueing network by a single queue such that the equilibrium distribution of the rest of the network remains unchanged.

[5] prove the aggregation method to be correct for queueing networks of the BCMP-type [2] consisting of two subnetworks of which the subnetwork of interest is a single station. This can easily be extended to subnetworks of interest consisting of several stations such that jobs enter the subnetwork through a single input node and leave the subnetwork through a single output node. [1], [11] and [16] further extend Norton's theorem to BCMP-networks consisting of two arbitrary subnetworks. Related work on decomposition and aggregation is reported in [8]. In [17] the aggregation results for BCMP-networks consisting of two subnetworks are generalized to queueing networks consisting of two quasireversible subnetworks. An approach related to the one based on Norton's theorem is reported in [6], [7], called "Parametric analysis by chain." This method constructs a reduced network around a particular routing chain of interest. From these results the general impression seems to have grown that aggregation and decomposition results are generally valid under product form conditions. However,

no formal justification for such results is available when state dependent routing and servicing such as due to blocking or common resource sharing are involved. The present paper aims to generalize these results to queueing networks with blocking and general state dependent service rates. We obtain the following theoretical results:

- Based on a decomposition of the state dependent routing probabilities into a global component and a local component we guarantee a decomposition of the routing part of the equilibrium distribution into a global component and a local component. These global and local components can be analyzed separately as if they were independent.
- Based on a decomposition of the state dependent service function into a local component and a global component, where “local component” and “global component” are determined by the decomposition of the routing probabilities, similarly we obtain a decomposition of the service part of the equilibrium distribution into a global component and a local component.
- By combination of the routing and servicing mechanism, the total detailed equilibrium distribution can be decomposed into an equilibrium distribution for global behaviour with aggregated transition rates and an equilibrium distribution for local behaviour with local transition rates. Moreover, the local equilibrium distribution factorizes into the subnetworks as if they are independent.
- By these results we have extended Norton’s theorem for queueing networks to state dependent routing and servicing mechanisms and aggregations to multiple components.
- A formal justification is provided to evaluate global characteristics by verifying global traffic equations and to evaluate local characteristics of parts of the queueing network by verifying global traffic equations and local traffic equations at the parts at interest of the queueing network only.

As practical implications of these results the following two methods can be proposed for analyzing the equilibrium distribution for queueing networks of which the routing and service mechanisms can be decomposed into local and global components:

(1) Monitoring method

1. Determine the global routing probabilities from the global structure of the queueing network.
2. Determine the global service characteristics by monitoring the state dependent output of the global parts of the queueing network.
3. Compute the global equilibrium distribution. This global distribution equals the equilibrium distribution of the original queueing network aggregated over subnetworks.

4. For the subnetworks for which the local distribution is of interest, first analyze the subnetwork in isolation with constant (e.g. unit) arrival rates. The equilibrium distribution of the subnetworks in isolation then equals the real equilibrium distribution of the subnetworks when incorporated in the network conditional on the global number of jobs present at the subnetwork.

(2) Computational method

1. Compute the equilibrium distribution of the subnetworks in isolation with constant (e.g. unit) arrival rates.
2. Calculate the global service characteristics from the equilibrium distribution of the subnetworks and the global part of the original service characteristics.
3. Compute the global equilibrium distribution from the global component of the routing probabilities and the global service characteristics.
4. Obtain the equilibrium distribution from the global and local equilibrium distributions.

The methods discussed above for analyzing the equilibrium distribution for queueing networks are motivated by Thevenin's rather than Norton's theorem from electrical circuit theory (cf. [3]). As will be discussed below, for electrical circuits Thevenin's theorem is essentially equivalent to Norton's theorem. The analog of Thevenin's theorem for queueing networks, however, is different and, as will be clarified in the paper, more general than the well-known analog of Norton's theorem. The differences in the structure of the analysis are the following (cf. Remarks 6.1, 6.4):

	Thevenin's theorem	Norton's theorem
1.	Subnetworks are analyzed as open queueing networks.	Subnetworks are analyzed as closed queueing networks.
2.	The service speed of aggregated subnetworks determines equilibrium distribution.	The throughput of a closed loop version of subnetworks determines equilibrium distribution.

- ad 1. Note that in the original non-aggregated queueing network the subnetworks appear as open queueing networks. Therefore, Thevenin's theorem is more natural in a queueing network environment than Norton's theorem.
- ad 2. The service speed of an aggregated subnetwork may depend on the total global (aggregated) state of the queueing network when Thevenin's theorem is applied. This is natural as Thevenin's theorem analyzes the global behaviour of the queueing network. For Norton's theorem, however, for obtaining the throughput of a closed loop version, the rest of the queueing network is ignored. Therefore, the throughput cannot depend on the state of the rest of the queueing network.

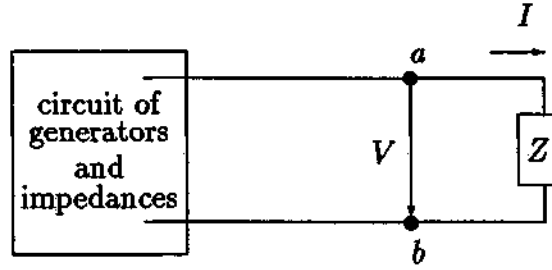


Figure 1. Electrical circuit

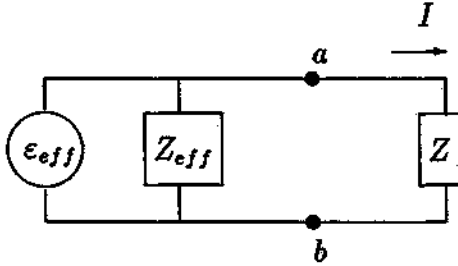
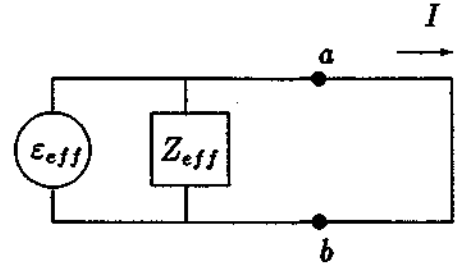


Figure 2. a: Equivalent electrical circuit



b: Z shorted

As Norton's theorem has become the standard phrasing in the literature on aggregation in queueing networks, the aggregation and decomposition results obtained here will be referred to as Norton's theorem.

To shed some light on these differences and to give some background information, let us discuss Thevenin's theorem and Norton's theorem for electrical circuits in some more detail. To this end, consider an electrical circuit consisting of generators and impedances as shown in Figure 1. Based on Kirchoff's laws the circuit of generators and impedances can be replaced by an effective generator and impedance. By Kirchoff's first law: *"The sum of the currents into any node must be zero."* we obtain Norton's theorem (cf. [3]) stating that the circuit of generators and impedances can be replaced by an effective generator and a parallel internal impedance without affecting the behaviour of Z (Figure 2a). The value of the current source is set equal to the current flowing from a to b when Z is replaced by a short (Figure 2b). By Kirchoff's second law: *"The sum of the potential differences around a complete loop of a circuit is equal to zero."* we obtain Thevenin's theorem (cf. [3]) stating that, without affecting the behaviour of Z , the circuit of generators and impedances can be replaced by an effective generator in series with an internal impedance (Figure 3a). The value of the potential difference generated by the generator is set equal to the potential difference between a and b when Z is removed (Figure 3b). The behaviour of Z in the equivalent network obtained via Norton's theorem and Thevenin's theorem is identical. The internal structure and behaviour of the replacement circuit consisting of ϵ_{eff} and Z_{eff} may differ. Obviously, analyzing the behaviour of Z in the equivalent network requires less computational effort.

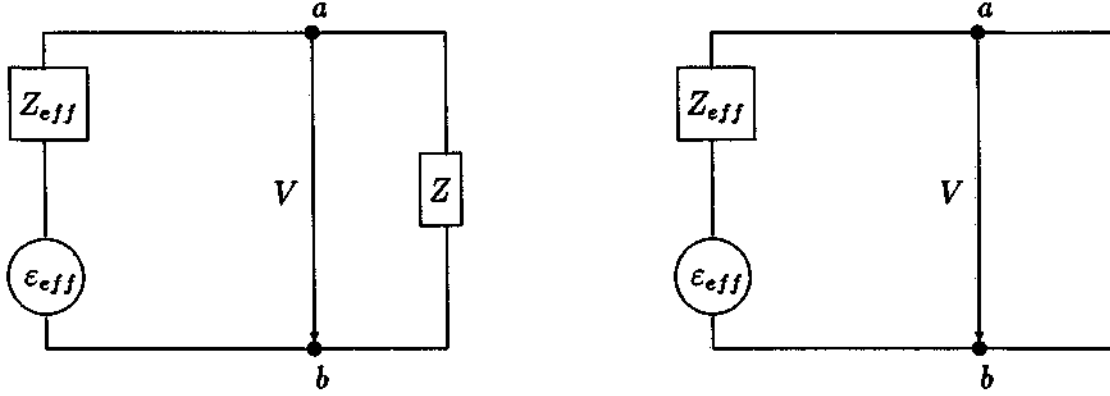


Figure 3. a: Equivalent electrical circuit

b: Z removed

This paper is organized as follows. Section 2 presents the general queueing network studied in this paper. We insist on using the highly theoretical form for the transition rates discussed in Section 2, but include a more practical example illustrating typical service functions included. In general the routing part of the transition rates is most difficult to analyze. Therefore, in Section 3 we first analyze the routing characteristics. Motivated by electrical circuit theory we decompose the routing probabilities into a global component and a local component and prove that this implies the same decomposition for the solution of the traffic equations. Second, the service characteristics are analyzed. Similar to the routing characteristics the service functions are decomposed into a global component and a local component. From this decomposition, in Section 4 we obtain a full decomposition of the equilibrium distribution resulting in Norton's theorem for queueing networks with state dependent routing and servicing characteristics. Section 5 gives some examples and Section 6 discusses the literature.

2 Model

Consider a queueing network consisting of N stations. A job at station i requires an exponential service with parameter μ_i , $i = 1, \dots, N$. Let $\bar{n} = (n_1, \dots, n_N)$ denote the number of jobs at the stations, i.e. n_i is the number of jobs at station i , $i = 1, \dots, N$. Let e_i denote the i -th unit vector, $i = 1, \dots, N$, and $e_0 = 0$. When the queueing network is in state $\bar{n}$ and a job at station i completes service it will route to station j with state dependent routing probability $p_{ij}(\bar{n})$, $j = 1, \dots, N$, and leaves the network with probability $p_{i0}(\bar{n}) = 1 - \sum_{j=1}^N p_{ij}(\bar{n})$. For notational convenience the outside is regarded as station 0, i.e. when a job leaves the network it routes to station 0 and when a job enters the network it routes from station 0. When the queueing network is in state $\bar{n}$ the service rate at station i is given by

$$\mu_i(\bar{n}) = \mu_i \frac{\psi(\bar{n} - e_i)}{\phi(\bar{n})}, \quad i = 0, \dots, N, \quad (2.1)$$

where ψ is a non-negative function and ϕ a positive function such that $\frac{\psi(\bar{n}-e_i)}{\phi(\bar{n})}$ represents the total service effort allocated to station i in state $\bar{n}$ and μ_0 is the parameter of the arrival process to the queueing network. The form (2.1) has been reported and is illustrated extensively in recent literature on product form queueing networks (cf. [4], [9], [13]). Roughly speaking, it is the most general service form possible to guarantee a product form equilibrium distribution. With service and routing functions specified above the queueing network can be represented as a continuous-time Markov chain at $N_0^N = \{0, 1, 2, \dots\}^N$ with transition rates $q(\bar{n}, \bar{n}')$, $\bar{n}, \bar{n}' \in N_0^N$, given by

$$q(\bar{n}, \bar{n}') = \begin{cases} \mu_i(\bar{n})p_{ij}(\bar{n}) & \bar{n}' = \bar{n} - e_i + e_j, i, j = 0, \dots, N, \\ 0 & \text{otherwise.} \end{cases} \quad (2.2)$$

Throughout this paper we will assume that the Markov chain is stable, regular, irreducible and that there exists a unique equilibrium distribution at state space $S \subset N_0^N$, i.e. a set of non-negative numbers $\pi = (\pi(\bar{n}), \bar{n} \in S)$, summing to unity, that satisfies the global balance equations at S (cf. [10], [19])

$$\pi(\bar{n}) \sum_{\bar{n}' \in S} q(\bar{n}, \bar{n}') = \sum_{\bar{n}' \in S} \pi(\bar{n}') q(\bar{n}', \bar{n}), \quad \bar{n} \in S.$$

As the transition rates have the special form (2.2), the global balance equations can be written

$$\pi(\bar{n}) \sum_{i=0}^N \mu_i(\bar{n}) = \sum_{i=0}^N \sum_{j=0}^N \pi(\bar{n} - e_i + e_j) \mu_j(\bar{n} - e_i + e_j) p_{ji}(\bar{n} - e_i + e_j), \quad \bar{n} \in S. \quad (2.3)$$

The following result expresses that product form type results can be concluded if the state dependent traffic equations per station can be solved. In its present form, which is tailored to the role of state dependent routing and blocking, it is directly adopted from [14], but various less explicit forms can be found in the literature (e.g. [4], [9], [13], [19]).

Result 2.1 *If a positive solution $H = (H(\bar{n}), \bar{n} \in S)$ exists to the state dependent traffic equations per station for all $\bar{n} \in S$, $i = 1, \dots, N$*

$$H(\bar{n}) = H(\bar{n} - e_i) \mu_0 p_{0i}(\bar{n} - e_i) + \sum_{j=1}^N H(\bar{n} - e_i + e_j) p_{ji}(\bar{n} - e_i + e_j), \quad (2.4)$$

then, with B a normalizing constant, the equilibrium distribution of the process with transition rates (2.2) is given by

$$\pi(\bar{n}) = B \phi(\bar{n}) H(\bar{n}) \prod_{k=1}^N \left(\frac{1}{\mu_k} \right)^{n_k}, \quad \bar{n} \in S. \quad (2.5)$$

Proof By substitution of (2.1) and (2.5) into (2.3), for each i separately (2.3) immediately reduces to (2.4). $\square$

Remark 2.2 (Product form) The equilibrium distribution (2.5) is said to be of product form in the sense that, except for the normalizing constant, it decomposes into a service part, $\phi(\bar{n}) \prod_{k=1}^N \left(\frac{1}{\mu_k}\right)^{n_k}$, and a routing part, $H(\bar{n})$. Moreover, the service part consists of a term representing the total state dependent service effort in the queueing network, $\phi(\bar{n})$, and a term representing the expected amount of service required at the stations, $\prod_{k=1}^N \left(\frac{1}{\mu_k}\right)^{n_k}$. This term, in turn, most characteristically for product form expressions factorizes into the individual stations. This decomposition into a service and routing part is complete in the sense that the service and routing part can be treated separately. This observation is used in Section 3 where we first consider the routing part of the equilibrium distribution and subsequently analyze the service part of the equilibrium distribution. $\square$

Remark 2.3 (Open and closed queueing networks; traffic equations) Our parameterization models both open and closed queueing networks. If the queueing network is closed, for all $\bar{n}$ and $i = 1, \dots, N$ we assume that $p_{i0}(\bar{n}) = p_{0i}(\bar{n}) = 0$. The traffic equations now read for $\bar{n} \in S$, $i = 1, \dots, N$

$$H(\bar{n}) = \sum_{j=1}^N H(\bar{n} - e_i + e_j) p_{ji}(\bar{n} - e_i + e_j).$$

If the queueing network is open, from (2.4) by summation we obtain the traffic equations for the outside:

$$H(\bar{n}) \mu_0 \sum_{j=1}^N p_{0j}(\bar{n}) = \sum_{j=1}^N H(\bar{n} + e_j) p_{j0}(\bar{n} + e_j).$$

$\square$

Before proceeding to the decomposition of the queueing network into subnetworks we first introduce some notation and give an example to illustrate that the service function $\mu_i(\bar{n})$ covers standard type service functions for Jacksonian networks.

Notation 2.4 (Clusters) In the sequel we will group the stations of the queueing network into clusters or subnetworks. A cluster consists of a number of stations such that clusters are disjoint, non-empty and all stations are part of a cluster, i.e. the stations are grouped into clusters C_r , $r \in R$ such that

$$\begin{aligned} C_r &\subset \{1, \dots, N\}, \quad r \in R, \\ C_r \cap C_s &= \emptyset, \quad r \neq s, \\ \bigcup_{r \in R} C_r &= \{1, \dots, N\}, \\ C_0 &= \text{outside.} \end{aligned}$$

Before proceeding to the decomposition of the queueing network, we define the following notation:

$$\begin{aligned} \bar{n}^{(r)} &= \bar{n}|_{C_r} && \text{number of jobs at the stations in } C_r, \\ e_i^{(r)} &= e_i|_{C_r} && \text{unit vector for station } i \text{ at } C_r, \\ N_r &= \sum_{i \in C_r} n_i && \text{total number of jobs at } C_r, \\ \bar{N} &= (N_1, \dots, N_R) \\ \bar{E}_r &&& \text{unit vector for cluster } r, \quad r \in R. \end{aligned}$$

Thus, for $\bar{n} \in S$ the vector $\bar{n}^{(r)}$ consists of the components of $\bar{n}$ inside C_r only and $\bar{N}$ gives the total number of jobs at the clusters. $\square$

Example 2.5 (Service function) The service function $\mu_i(\bar{n})$ as given in (2.1) is the most general form appearing in the literature and is of theoretical interest as it combines generality with theoretical elegance. An extensive illustration of the service rates that can be modeled can be found in [13] and in batch movement setups in [4], [9]. In practical cases, for a given function $\mu_i(\bar{n})$, the difficult task is to find functions ψ and ϕ such that the service function can be written in the form (2.1). A natural and illustrative example of a service function that can be written in the form (2.1) is the following. Assume that the service speed at station i is a function of the number of jobs at station i , $f_i(n_i)$ if n_i jobs are present at station i , and the number of jobs at the cluster containing station i , say cluster r , $F_r(N_r)$ if N_r jobs are present at cluster r . This gives

$$\mu_i(\bar{n}) = \mu_i f_i(n_i) F_r(N_r), \quad i \in C_r,$$

which, for ψ and ϕ given by

$$\psi(\bar{n}) = \phi(\bar{n}) = \left[\prod_{i=1}^N \prod_{k=1}^{n_i} f_i(k) \right]^{-1} \left[\prod_{r \in R} \prod_{k=1}^{N_r} F_r(k) \right]^{-1}$$

can be written in the form (2.1) by noting that $\frac{\phi(\bar{n}-e_i)}{\phi(\bar{n})} = f_i(n_i) F_r(N_r)$. $\square$

Remark 2.6 (Service function) For a service function specified in advance, some freedom remains in determining the terms appearing in $\mu_i(\bar{n})$. Firstly, ψ and ϕ are determined up to a common constant. Without loss of generality, this constant can be fixed by assuming $\phi(\bar{0}) = 1$. Secondly, μ_i and ψ are determined up to a common constant. Without loss of generality, this constant can be fixed by assigning a fixed value to the arrival parameter μ_0 . When the queueing network is analyzed for a fixed parameter μ_0 , we obtain the service characteristics for μ'_0 by substituting

$$\mu'_i = \mu_i \frac{\mu'_0}{\mu_0}, \quad i = 0, \dots, N, \quad \psi'(\bar{n}) = \frac{\mu_0}{\mu'_0} \psi(\bar{n}),$$

which leaves the service function $\mu_i(\bar{n})$ unchanged. Also note that the equilibrium distribution is not affected by this substitution as H is proportional to μ_0 too. If H is a solution to the traffic equations (2.4) for μ_0 , then H' given by

$$H'(\bar{n}) = H(\bar{n}) \prod_{k=1}^N \left(\frac{\mu'_0}{\mu_0} \right)^{n_k}$$

is a solution to the traffic equations for μ'_0 . This observation will be used in Example 5.1. $\square$

3 Decomposition into clusters

As is discussed in Remark 2.2, the routing part and the service part of the transition rates can be analyzed separately. In general the routing part is more complex than the service part. Therefore, in Section 3.1 we first consider the routing part of the transition rates. In Section 3.3 we will consider the service part. Examples illustrating the decomposition of the routing and service parts are given in Sections 3.2, 3.4.

3.1 Routing: Conditions

The aim of this section is to decompose the routing characteristics $p_{ij}(\cdot)$ and $H(\cdot)$ into a global part and a local part. To this end, by analogy with the structure of electrical circuits at which Thevenin's theorem can be applied (cf. [3]) we decompose the routing probabilities into a global component and a local component (Assumption 3.1). This natural decomposition immediately implies a full decomposition of H into a global part and a local part. Moreover, the global part of H is completely determined by the global routing probabilities and the local part of H is completely determined by the local routing probabilities.

The key observation for electrical circuits at which Thevenin's theorem can be applied is that an electrical circuit consisting of generators and impedances can be replaced by an effective generator and impedance if and only if this circuit has a single input node (e.g. b in Figure 1) and a single output node (e.g. a in Figure 1). As the behaviour of queues is in general much more complex than the behaviour of impedances, which more or less resemble $M|M|\infty$ queues, for a similar theorem to hold in a queueing network the routing from one cluster of stations to another must be independent of the actual stations within these clusters. The following assumption guarantees this behaviour for the queueing network discussed in Section 2.

Assumption 3.1 (Routing probabilities) *Assume that for $i, j = 0, \dots, N$ and $\bar{n} \in S$ the routing probabilities have the form*

$$p_{ij}(\bar{n}) = \begin{cases} p_{ij}^{(r)}(\bar{n}^{(r)}) + p_{i0}^{(r)}(\bar{n}^{(r)})p^{rr}(\bar{N})p_{0j}^{(r)}(\bar{n}^{(r)} - e_i^{(r)}), & i, j \in C_r, \\ p_{i0}^{(r)}(\bar{n}^{(r)})p^{rs}(\bar{N})p_{0j}^{(s)}(\bar{n}^{(s)}), & i \in C_r, j \in C_s, \\ p_{i0}^{(r)}(\bar{n}^{(r)})p^{r0}(\bar{N}), & i \in C_r, j \in C_0, \\ p^{0s}(\bar{N})p_{0j}^{(s)}(\bar{n}^{(s)}), & i \in C_0, j \in C_s, \end{cases} \quad (3.1)$$

where $p_{ij}^{(r)}$, $i, j \in C_r \cup \{0\}$, $r \in R$, and p^{rs} , $r, s \in R \cup \{0\}$, satisfy for all $\bar{n}$

$$\begin{aligned} \sum_{j \in C_s} p_{0j}^{(s)}(\bar{n}^{(s)}) &= 1, \quad \sum_{j \in C_s} p_{ij}^{(s)}(\bar{n}^{(s)}) + p_{i0}^{(s)}(\bar{n}^{(s)}) = 1, \\ \sum_{s \in R} p^{0s}(\bar{N}) &= 1, \quad \sum_{s \in R} p^{rs}(\bar{N}) + p^{r0}(\bar{N}) = 1. \end{aligned}$$

Remark 3.2 (Routing probabilities) Assumption 3.1 implies that for the input and output into a cluster the behaviour of a cluster in the queueing network with routing probabilities p_{ij} is similar to the structure of an electrical circuit at which Thevenin's theorem can be applied. To show this, consider a job routing from station i at cluster r to station j at cluster s . First, according to $p_{i0}^{(r)}$ the job leaves queue i and routes to an input/output node for cluster r (e.g. b in Figure 1). Subsequently, the job is routed from the input/output node of cluster r to the input/output node of cluster s (e.g. a in Figure 1) according to p^{rs} independent of the local state of the clusters. Finally, the job is routed from the input/output node of cluster s to queue j at cluster s according to $p_{0j}^{(s)}$.

$p_{ij}^{(r)}$ depends on $\bar{n}^{(r)}$, the local state of cluster r , only, and p^{rs} depends on $\bar{N}$, the global state of the network, only. Therefore, the routing probabilities $p_{ij}^{(r)}$ can be interpreted as the state dependent routing probabilities at cluster r when cluster r would be considered in isolation, i.e. when cluster r is considered as a queueing network without interaction with the rest of the network. In contrast, p^{rs} can be interpreted as the state dependent routing probabilities for the network in which each cluster is replaced by a single queue. This interpretation will be formalized in Section 4, when the transition rates of the queueing network at local level and at global level are defined. $\square$

Remark 3.3 (Blocking) Blocking can arise at local level due to $p_{ij}^{(r)}(\bar{n}^{(r)})$ or at global level due to $p^{rs}(\bar{N})$. These types of blocking are reflected in the functions $H^{(r)}$ and H_R below and are completely independent. Note that jobs arriving at a cluster may be blocked since $p_{0j}^{(r)}(\bar{n}^{(r)}) = 0$ is allowed. However, a job arriving at a cluster cannot be rejected due to $p_{0j}^{(r)}(\cdot)$. Therefore, if $p_{0j}^{(r)}(\bar{n}^{(r)}) = 0$ for some j an arriving job must route to another station of the cluster. $\square$

As the routing probabilities are separated into a global component and a local component which can be interpreted as the global and local routing probabilities, it is natural to investigate when the traffic equations (2.4) can be decomposed into global and local traffic equations. We therefore present the following conditions.

Condition 3.4 (Global traffic equations) *There exists a positive solution, H_R , to the global traffic equations per cluster. That is, for each subnetwork $r \in R$ and for all $\bar{N}$ such that $\bar{n} \in S$:*

$$H_R(\bar{N}) = \sum_{s \in R} H_R(\bar{N} - \bar{e}_r + \bar{e}_s) p^{sr}(\bar{N} - \bar{e}_r + \bar{e}_s) + \mu_0 H_R(\bar{N} - \bar{e}_r) p^{0r}(\bar{N} - \bar{e}_r). \quad (3.2)$$

Condition 3.5 (Local traffic equations) *For $r \in R$ there exists a positive solution, $H^{(r)}$, to the local traffic equations per station. That is, for each $i \in C_r$ and all $\bar{n}^{(r)}$ such that $\bar{n} \in S$:*

$$\begin{aligned} H^{(r)}(\bar{n}^{(r)}) &= \sum_{j \in C_r} H^{(r)}(\bar{n}^{(r)} - e_i^{(r)} + e_j^{(r)}) p_{ji}^{(r)}(\bar{n}^{(r)} - e_i^{(r)} + e_j^{(r)}) \\ &\quad + H^{(r)}(\bar{n}^{(r)} - e_i^{(r)}) p_{0i}^{(r)}(\bar{n}^{(r)} - e_i^{(r)}). \end{aligned} \quad (3.3)$$

From Assumption 3.1 we obtain a decomposition of the routing probabilities into a part characterizing the global behaviour and a part characterizing the local behaviour. For H to be decomposable into a global and local part compatible with the assumption and conditions stated above, $H(\bar{n})$ may be expected to factorize into a part involving $H^{(r)}$ and a part involving H_R . Let us investigate this in more detail. Insertion of (3.1) into (2.4) gives, for $i \in C_r$, $r \in R$, $\bar{n} \in S$

$$\begin{aligned}
H(\bar{n}) = & \sum_{j \in C_r} H(\bar{n} - e_i + e_j) p_{ji}^{(r)} (\bar{n}^{(r)} - e_i^{(r)} + e_j^{(r)}) \\
& + \left\{ \sum_{j \in C_r} H(\bar{n} - e_i + e_j) p_{j0}^{(r)} (\bar{n}^{(r)} - e_i^{(r)} + e_j^{(r)}) p^{rr}(\bar{N}) \right. \\
& + \sum_{s \in R, s \neq r} \sum_{j \in C_s} H(\bar{n} - e_i + e_j) p_{j0}^{(s)} (\bar{n}^{(s)} + e_j^{(s)}) p^{sr}(\bar{N} - \bar{E}_r + \bar{E}_s) \\
& \left. + \mu_0 H(\bar{n} - e_i) p^{0r}(\bar{N} - \bar{E}_r) \right\} p_{0i}^{(r)} (\bar{n}^{(r)} - e_i^{(r)}). \quad (3.4)
\end{aligned}$$

In this form the traffic equations show great similarities to the traffic equations at local level (3.3), when the part inside brackets $\{ \}$ is replaced by $H(\bar{n} - e_i)$ and although not as obvious, the traffic equations at global level (3.2). This strongly suggests that H will indeed factorize into a global and a local component. This is established in the following theorem.

Theorem 3.6 (Routing decomposition) *Assume that H_R is a solution to the global traffic equations, (3.2), and $H^{(r)}$ is a solution to the local traffic equations, (3.3). Then $H(\bar{n})$ defined as*

$$H(\bar{n}) = H_R(\bar{N}) \prod_{r \in R} H^{(r)}(\bar{n}^{(r)}), \quad \bar{n} \in S, \quad (3.5)$$

is a solution of the traffic equations (2.4).

Proof Summation of (3.2) gives that if $H^{(r)}$ satisfies (3.3) then for all $\bar{n}^{(r)}$ such that $\bar{n} \in S$ $H^{(r)}$ also satisfies

$$H^{(r)}(\bar{n}^{(r)}) = \sum_{j \in C_r} H^{(r)}(\bar{n}^{(r)} + e_j^{(r)}) p_{j0}^{(r)} (\bar{n}^{(r)} + e_j^{(r)}). \quad (3.6)$$

Consider the right-hand side of the traffic equations (3.4). For notational convenience, we have divided (3.4) by $H(\bar{n} - e_i)$. This is not essential to the proof. By insertion of (3.5) and from (3.6), (3.2), (3.3) we obtain for $i \in C_r$, $\bar{n} \in S$

$$\begin{aligned}
& \sum_{j \in C_r} \frac{H(\bar{n} - e_i + e_j)}{H(\bar{n} - e_i)} p_{ji}^{(r)} (\bar{n}^{(r)} - e_i^{(r)} + e_j^{(r)}) \\
& + \left\{ \sum_{j \in C_r} \frac{H(\bar{n} - e_i + e_j)}{H(\bar{n} - e_i)} p_{j0}^{(r)} (\bar{n}^{(r)} - e_i^{(r)} + e_j^{(r)}) p^{rr}(\bar{N}) \right.
\end{aligned}$$

$$\begin{aligned}
& + \sum_{s \in R, s \neq r} \sum_{j \in C_s} \frac{H(\bar{n} - e_i + e_j)}{H(\bar{n} - e_i)} p_{j0}^{(s)}(\bar{n}^{(s)} + e_j^{(s)}) p^{sr}(\bar{N} - \bar{E}_r + \bar{E}_s) \\
& + \mu_0 H(\bar{n} - e_i) p^{0r}(\bar{N} - \bar{E}_r) \left\} p_{0i}^{(r)}(\bar{n}^{(r)} - e_i^{(r)}) \right. \\
& \stackrel{(3.5)}{=} \sum_{j \in C_r} \frac{H_R(\bar{N})}{H_R(\bar{N} - \bar{E}_r)} \frac{H^{(r)}(\bar{n}^{(r)} - e_i^{(r)} + e_j^{(r)})}{H^{(r)}(\bar{n}^{(r)} - e_i^{(r)})} p_{ji}^{(r)}(\bar{n}^{(r)} - e_i^{(r)} + e_j^{(r)}) \\
& + \left\{ \sum_{j \in C_r} \frac{H_R(\bar{N})}{H_R(\bar{N} - \bar{E}_r)} \frac{H^{(r)}(\bar{n}^{(r)} - e_i^{(r)} + e_j^{(r)})}{H^{(r)}(\bar{n}^{(r)} - e_i^{(r)})} p_{j0}^{(r)}(\bar{n}^{(r)} - e_i^{(r)} + e_j^{(r)}) p^{rr}(\bar{N}) \right. \\
& + \sum_{s \in R, s \neq r} \sum_{j \in C_s} \frac{H_R(\bar{N} - \bar{E}_r + \bar{E}_s)}{H_R(\bar{N} - \bar{E}_r)} \frac{H^{(s)}(\bar{n}^{(s)} + e_j^{(s)})}{H^{(s)}(\bar{n}^{(s)})} p_{j0}^{(s)}(\bar{n}^{(s)} + e_j^{(s)}) p^{sr}(\bar{N} - \bar{E}_r + \bar{E}_s) \\
& \left. + \mu_0 p^{0r}(\bar{N} - \bar{E}_r) \right\} p_{0i}^{(r)}(\bar{n}^{(r)} - e_i^{(r)}) \\
& \stackrel{(3.6)}{=} \sum_{j \in C_r} \frac{H_R(\bar{N})}{H_R(\bar{N} - \bar{E}_r)} \frac{H^{(r)}(\bar{n}^{(r)} - e_i^{(r)} + e_j^{(r)})}{H^{(r)}(\bar{n}^{(r)} - e_i^{(r)})} p_{ji}^{(r)}(\bar{n}^{(r)} - e_i^{(r)} + e_j^{(r)}) \\
& + \left\{ \sum_{s \in R} \frac{H_R(\bar{N} - \bar{E}_r + \bar{E}_s)}{H_R(\bar{N} - \bar{E}_r)} p^{sr}(\bar{N} - \bar{E}_r + \bar{E}_s) + \mu_0 p^{0r}(\bar{N} - \bar{E}_r) \right\} p_{0i}^{(r)}(\bar{n}^{(r)} - e_i^{(r)}) \\
& \stackrel{(3.2)}{=} \sum_{j \in C_r} \frac{H_R(\bar{N})}{H_R(\bar{N} - \bar{E}_r)} \frac{H^{(r)}(\bar{n}^{(r)} - e_i^{(r)} + e_j^{(r)})}{H^{(r)}(\bar{n}^{(r)} - e_i^{(r)})} p_{ji}^{(r)}(\bar{n}^{(r)} - e_i^{(r)} + e_j^{(r)}) \\
& + \frac{H_R(\bar{N})}{H_R(\bar{N} - \bar{E}_r)} p_{0i}^{(r)}(\bar{n}^{(r)} - e_i^{(r)}) \\
& \stackrel{(3.3)}{=} \frac{H_R(\bar{N})}{H_R(\bar{N} - \bar{E}_r)} \frac{H^{(r)}(\bar{n}^{(r)})}{H^{(r)}(\bar{n}^{(r)} - e_i^{(r)})} \stackrel{(3.5)}{=} \frac{H(\bar{n})}{H(\bar{n} - e_i)}. \quad \square
\end{aligned}$$

Remark 3.7 (Reference [14]) A decomposition of the routing probabilities similar to (3.1) is introduced in [14]. This reference concentrates on the notion of blocking at global level and on the characterization of product form global blocking structures. More precisely, it purely concentrates on the global traffic equations (3.2). The local routing probabilities were assumed to be state independent. Most notably, a decomposition result of the form (3.5) into a global and local solution was not concluded. In this respect, the present paper generalizes the results obtained in [14]. $\square$

3.2 Routing: Examples

The following examples illustrate the decomposition of the routing characteristics into a global and a local component. Example 3.8 illustrates the decomposition in the case of state independent routing. This example also shows that, at least for the case of state independent routing, a decomposition of H into global and local components satisfying the global and local traffic equations is possible if a solution exists to the standard traffic equations. This justifies Conditions 3.4, 3.5. Example 3.9 is adopted from [14] and considers state dependent global routing. In this example a capacity constraint is imposed on a subnetwork, say no more than U_2 jobs are allowed to be present at cluster 2 simultaneously. To retain a solution to the global traffic equations various blocking protocols are allowed. We will illustrate the global recirculation protocol stating that at all non-saturated clusters departing jobs are rerouted into the cluster. Finally, Example 3.10 shows that the decomposition of the routing characteristics into a global and a local component must be based on the structure of the queueing network. Most notably, blocking effects play a crucial role in this decomposition.

Example 3.8 (State independent routing) Assume that the routing probabilities are state independent, i.e.

$$p_{ij}^{(r)}(\bar{n}^{(r)}) = p_{ij}^{(r)}, \quad i, j \in C_r \cup \{0\}, \quad p^{rs}(\bar{n}) = p^{rs}, \quad r, s \in R \cup \{0\}.$$

Assume that $\{a^{(r)}\}_{r \in R}$ is a solution of the global traffic equations

$$a^{(r)} = \sum_{s \in R} a^{(s)} p^{sr} + \mu_0 p^{0r}, \quad r \in R, \quad (3.7)$$

and for each r that $\{c_i^{(r)}\}_{i \in C_r}$ is a solution of the local traffic equations

$$c_i^{(r)} = \sum_{j \in C_r} c_j p_{ji}^{(r)} + p_{0i}^{(r)}, \quad i \in C_r. \quad (3.8)$$

From these assumptions we obtain that Conditions 3.4, 3.5 are satisfied as

$$H_R(\bar{n}) = \prod_{r \in R} (a^{(r)})^{N_r} \quad \text{and} \quad H^{(r)}(\bar{n}^{(r)}) = \prod_{j \in C_r} (c_j^{(r)})^{n_j^{(r)}} \quad (3.9)$$

satisfy (3.2) and (3.3) respectively. Furthermore, we immediately obtain that $\{c_i\}_{i=1}^N$ given by

$$c_i = c_i^{(r)} a^{(r)}, \quad i \in C_r, \quad (3.10)$$

is a solution to the standard traffic equations

$$c_i = \sum_{j=1}^N c_j p_{ji} + \mu_0 p_{0i}, \quad i = 1, \dots, N. \quad (3.11)$$

This implies that the solution H to (2.4) indeed satisfies (3.5) as from (3.9) and (3.10) we obtain

$$H(\bar{n}) = \prod_{k=1}^N c_k^{n_k} = \prod_{r \in R} (a^{(r)})^{N_r} \prod_{j \in C_r} (c_j^{(r)})^{n_j^{(r)}} = H_R(\bar{N}) \prod_{r \in R} H^{(r)}(\bar{n}^{(r)}).$$

Furthermore, this example illustrates that the total traffic equations (3.11) can be solved by solving a number of subproblems (3.8) and a global problem (3.7).

The reasoning above can be reversed to show that Conditions 3.4, 3.5 are satisfied if a solution exists to the traffic equations (3.11). To this end, note that if $\{c_i\}_{i=1}^N$ is a solution to (3.11) then $\{a^{(r)}\}_{r \in R}$ defined as

$$a^{(r)} = \sum_{j \in C_r} c_j p_{j0}^{(r)}, \quad r \in R,$$

is a solution to the global traffic equations (3.7), and $\{c_i^{(r)}\}_{i \in C_r, r \in R}$ defined as

$$c_i^{(r)} = \frac{c_i}{\sum_{j \in C_r} c_j p_{j0}^{(r)}}, \quad i \in C_r,$$

is a solution to the local traffic equations (3.8) as can be seen by substitution. $\square$

Example 3.9 (Conservative blocking; global recirculation protocol) Consider the following modification of the closed version of Example 3.8 (cf. [14]). A capacity constraint is imposed on cluster r , stating that no more than U_r jobs are allowed to be present at cluster r simultaneously. As arrivals into cluster r are blocked when this cluster is saturated, i.e. when $N_r = U_r$, the global routing probabilities must be modified. To retain a solution to the global traffic equations (3.2), a modification of the global routing probabilities is obtained from the global recirculation protocol. This protocol states that departures from all other clusters $s \neq r$ and arrivals into the queueing network are prohibited, which is achieved by letting jobs departing from cluster $s \neq r$ recirculate into cluster s while jobs arriving to the network are blocked. As a consequence, no more than one cluster can become saturated at the same time, which explains the name conservative blocking. (For an extensive discussion of global blocking phenomena the reader is referred to [14].) For the global recirculation protocol, the global routing probabilities take the form

$$p^{rs}(\bar{N}) = \begin{cases} p^{rs} & \text{if } N_k < U_k \text{ for all } k \neq r, \\ 1 & \text{if } N_k = U_k \text{ for some } k \neq r \text{ for } r = s. \end{cases}$$

As a consequence, with C the set of admissible states,

$$C = \{\bar{N} | N_r \leq U_r, \text{ for all } r \in R, \text{ and } N_r + N_s < U_r + U_s \text{ for any } r \neq s\},$$

and $1(A)$ the indicator of event A , i.e. $1(A) = 1$ if A is satisfied and $1(A) = 0$ otherwise, the global traffic equations require for each subnetwork $r \in R$, for $\bar{N} \in C$:

$$\begin{aligned} H_R(\bar{N}) &= \sum_{s \in R, s \neq r} H_R(\bar{N} - \bar{E}_r + \bar{E}_s) 1(\bar{N} - \bar{E}_r + \bar{E}_s \in C) 1(N_k < U_k \text{ for all } k \neq s) p^{sr} \\ &\quad + H_R(\bar{N}) 1(N_k = U_k \text{ for some } k \neq r). \end{aligned} \quad (3.12)$$

Observing that $N_k = U_k$ for some $k \neq r$, say $N_q = U_q$, implies that $1(N_k < U_k \text{ for all } k \neq s) = 1$ iff $s = q$ gives for all $s \neq r$ that $\bar{N} - \bar{E}_r + \bar{E}_s \notin C$ if $N_k = U_k$ for some $k \neq r$ which immediately gives that H_R given in Example 3.8 is a solution to (3.12). As the local routing characteristics are not affected by this modification, (3.9) remains valid.

This example shows that blocking at global level is possible, while retaining a solution of the global traffic equations and without affecting the local routing characteristics, $p_{ij}^{(r)}$ and $H^{(r)}$. Note that blocking at global level does affect the actual inputs of the subnetworks. This input process, however, is a global process. As subnetworks are analyzed with constant (unit) input rate, changing the global input process to the subnetworks does not affect the local solution $H^{(r)}$. Changing the input process to the subnetworks is a scaling effect for H and is incorporated into H_R . Let us illustrate

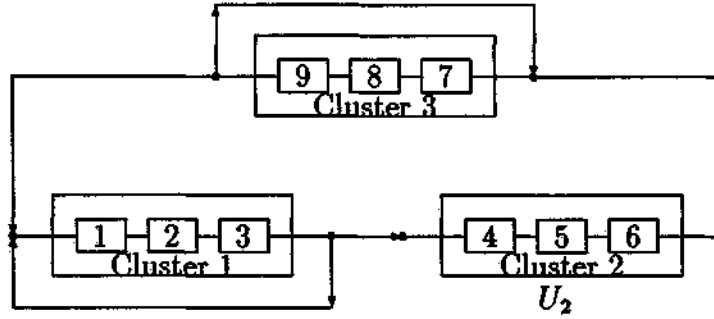


Figure 4. Capacity constraint U_2 at cluster 2

this with a concrete example. To this end, for simplicity, consider the cyclic queueing network consisting of 9 stations grouped into 3 clusters as depicted in Figure 4. A capacity constraint is imposed on cluster 2, stating that the total number of jobs present at cluster 2 cannot exceed U_2 . As a consequence of the global recirculation protocol, when cluster 2 is saturated departures from stations 3 and 9 are recirculated into stations 1 and 7 respectively. As the routing probabilities are modified by the global recirculation protocol, the traffic equations change too as is shown in the table below.

	System without blocking	System with blocking
Traffic equations (2.4)		
1	$H(\bar{n}) = H(\bar{n} - e_i + e_{i+1}), \forall i, \forall \bar{n}$	$H(\bar{n}) = H(\bar{n} - e_i + e_{i+1}), i \neq 1, 7, \forall \bar{n}$
2		$H(\bar{n}) = H(\bar{n} - e_i + e_{i+1}), i = 1, 7, N_2 < U_2$
3		$H(\bar{n}) = H(\bar{n} - e_i + e_{i+2}), i = 1, 7, N_2 = U_2$
Global equations (3.2)		
4	$H(\bar{N}) = H(\bar{N} - \bar{E}_r + \bar{E}_{r+1}), \forall r, \forall \bar{N}$	$H(\bar{N}) = H(\bar{N} - \bar{E}_r + \bar{E}_{r+1}), \forall r, N_2 < U_2$
5		$H(\bar{N}) = H(\bar{N} - \bar{E}_r + \bar{E}_{r+1}), r = 2, N_2 = U_2$
6		$H(\bar{N}) = H(\bar{N}), r = 1, 3, N_2 = U_2$
Local equations (3.3)		
7	$H^{(r)}(\bar{n}^{(r)}) = H^{(r)}(\bar{n}^{(r)} + e_i^{(r)}), i = 1, 4, 7$	$H^{(r)}(\bar{n}^{(r)}) = H^{(r)}(\bar{n}^{(r)} + e_i^{(r)}), i = 1, 4, 7$
8	$H^{(r)}(\bar{n}^{(r)}) = H^{(r)}(\bar{n}^{(r)} - e_i^{(r)} + e_{i+1}^{(r)}),$ $i = 2, 3, 5, 6, 8, 9$	$H^{(r)}(\bar{n}^{(r)}) = H^{(r)}(\bar{n}^{(r)} - e_i^{(r)} + e_{i+1}^{(r)}),$ $i = 2, 3, 5, 6, 8, 9$

Note that the traffic equations (2.4) are modified at (3) only. When cluster 2 is non-saturated a positive outrate of station 7 is balanced by a positive inrate from station 6. However, when cluster 2 is saturated this outrate is balanced by a positive inrate from station 9. This modification is due to the global recirculation protocol and is due to the modification of $p^*(\bar{N})$ only. This can immediately be seen from the global and local traffic equations. As the subnetworks in isolation are analyzed using a state dependent input process with unit rate relation (7) and (8) are not affected by global blocking phenomena. The global traffic equations, however, are modified if $N_2 = U_2$. As $\sum_{j \in C_r} n_j - (e_i)_j + (e_{i+1})_j = N_r = \sum_{j \in C_r} n_j$, (6) completely characterizes the modification (3). $\square$

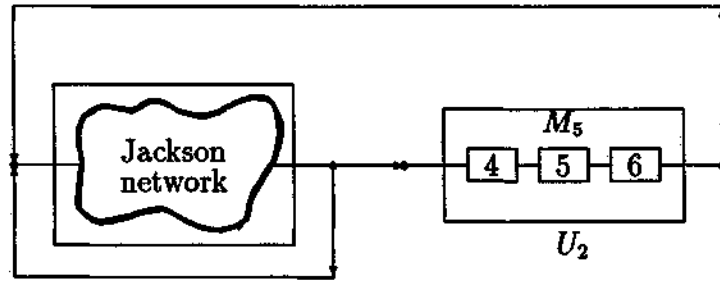


Figure 5. Capacity constraints M_5 at station 5 and U_2 at cluster 2

Example 3.10 (Choice of clusters) In Assumption 3.1 and in Examples 3.8, 3.9 the decomposition of the routing probabilities is imposed on the queueing network. In practical applications, however, this decomposition must be established based on the structure of the queueing network. This example shows that blocking effects play a crucial role in this decomposition. To this end, consider the queueing network consisting of a Jackson subnetwork and a tandem subnetwork as illustrated in Figure 5. Assume a capacity constraint, U_2 , is imposed on the total number of jobs at stations 4, 5 and 6 simultaneously and in addition the number of jobs at station 5 is constrained not to exceed M_5 . If we are interested in the global behaviour of the Jackson subnetwork and the tandem subnetwork, due to the capacity constraint, U_2 , a natural decomposition, similar to Example 3.9, would be to aggregate both the Jackson subnetwork and the tandem subnetwork, where, as blocking is involved, the recirculation protocol is used to modify the routing probabilities. However, the capacity constraint at station 5 affects the global routing. This can be seen as follows. If station 5 is saturated, jobs cannot route from station 4 to station 5. Therefore, according to the recirculation protocol, these jobs are rerouted into station 4. Similarly, jobs leaving the Jackson subnetwork are rerouted. Therefore, the global routing is affected by the state of station 5. As this is excluded by Assumption 3.1, the decomposition into the Jackson subnetwork and the tandem subnetwork cannot be modeled. However, the queueing network can be decomposed into the Jackson cluster and stations 4, 5 and 6 seen as clusters 4, 5 and 6, i.e. the tandem cluster cannot be aggregated, but the Jackson cluster can be aggregated into a single queue. $\square$

3.3 Servicing: Conditions

In Section 3.1 the solution to the traffic equations is decomposed into a global and a local component due to a similar decomposition of the routing probabilities. This section aims to obtain similar decomposition results for the servicing mechanism. As a first consequence, the equilibrium distribution will then factorize into global and local components.

From the natural decomposition of the routing probabilities (3.1) in Section 3, we have seen that the solution to the traffic equations consists of a global part describing the global routing characteristics and a local part describing the routing characteristics at the subnetworks. To obtain a similar result for the servicing part we also make the following natural assumption.

Assumption 3.11 (Service function ϕ) *Assume that the service function, ϕ , has the form*

$$\phi(\bar{n}) = \phi_R(\bar{N}) \prod_{r \in R} \phi^{(r)}(\bar{n}^{(r)}), \quad (3.13)$$

where $\phi_R(\bar{0}) = 1$ and $\phi^{(r)}(\bar{0}^{(r)}) = 1$, $r \in R$.

By (3.5) and (3.13) the equilibrium distribution (2.5) decomposes into global and local routing and servicing components:

$$\pi(\bar{n}) = B \phi_R(\bar{N}) H_R(\bar{N}) \prod_{r \in R} \phi^{(r)}(\bar{n}^{(r)}) H^{(r)}(\bar{n}^{(r)}) \prod_{k \in C_r} \left(\frac{1}{\mu_k} \right)^{n_k^{(r)}}, \quad \bar{n} \in S. \quad (3.14)$$

Assumption 3.11 establishes a decomposition of the equilibrium distribution into a global and a local part. In this case, it is natural that the total servicing mechanism decomposes into a global and local part also. To this end we also include the following assumption.

Assumption 3.12 (Service function ψ) *Assume that the service function, ψ , has the form*

$$\psi(\bar{n}) = \psi_R(\bar{N}) \prod_{r \in R} \psi^{(r)}(\bar{n}^{(r)}). \quad (3.15)$$

From Assumptions 3.11, 3.12 we now obtain that the service rate (2.1) decomposes into a global and a local part:

$$\mu_i(\bar{n}) = \mu_i \frac{\psi_R(\bar{N} - E_s)}{\psi_R(\bar{N})} \prod_{r \in R} \frac{\psi^{(r)}(\bar{n}^{(r)} - e_i^{(r)})}{\psi^{(r)}(\bar{n}^{(r)})}, \quad i \in C_s.$$

This decomposition is studied in more detail in Section 4.

3.4 Servicing: Examples

The following examples illustrate the service rates included in the paper and the decomposition into global and local service characteristics. In Example 3.13 the practical service rates described in Example 2.5 are decomposed into a global and a local component. Examples 3.14 and 3.15 illustrate that global state dependent service rates allow us to modify the service speed at clusters so as to avoid congestion.

Example 3.13 (Factorizing form) The service function given in Example 2.5 factorizes into a global and a local component. To this end, note that ϕ can be written in the form (3.13) with

$$\phi_R(\bar{N}) = \left[\prod_{r \in R} \prod_{k=1}^{N_r} F_r(k) \right]^{-1}, \quad \phi^{(r)}(\bar{n}^{(r)}) = \left[\prod_{j \in C_r} \prod_{k=1}^{n_j} f_j(k) \right]^{-1}. \quad (3.16)$$

Example 3.14 (Service speed reduction to help avoid congestion) This example shows that due to the total global state dependency of the service functions ψ and ϕ the service rate at the clusters can be chosen such that it helps to avoid congestion. To this end, reconsider the service rates introduced in Example 2.5:

$$\mu_i(\bar{n}) = \mu_i f_i(n_i) F_r(N_r), \quad i \in C_r. \quad (3.17)$$

In addition, assume that the number of jobs at cluster 2 is constrained not to exceed U_2 . As we have seen in Example 3.9 as soon as the total number of jobs at cluster 2 reaches its upper bound jobs leaving the other clusters are recirculated to be processed by the same cluster once more. As a more smooth alternative we can slow down all clusters $s \neq 2$ before cluster 2 reaches its upper bound. For example, we may reduce the service speed at clusters $s \neq 2$ by $\frac{1}{2}$ if the number of jobs at cluster 2 exceeds some given value L_2 . As we usually cannot alter the input process, this process remains unchanged. This gives

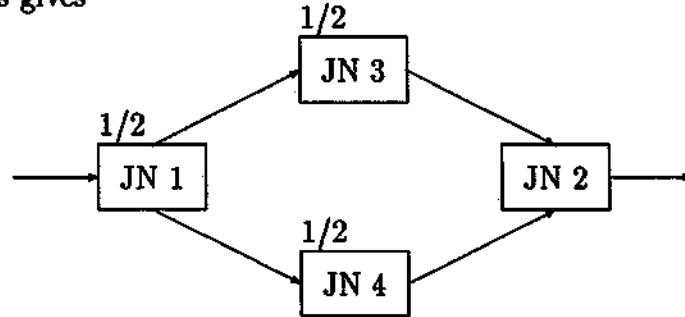


Figure 6. Service speed reduction

$$\hat{\mu}_i(\bar{n}) = \begin{cases} \mu_i f_i(n_i) F_r(N_r), & i \in C_r, r = 2, \\ \mu_i f_i(n_i) F_r(N_r), & i \in C_r, r \neq 2, N_2 < L_2, \\ \frac{1}{2} \mu_i f_i(n_i) F_r(N_r), & i \in C_r, r \neq 2, N_2 \geq L_2, \\ \mu_0, & i = 0. \end{cases}$$

This service rate can be written in the form (2.1). To this end, with ϕ_R given in (3.16), define

$$\hat{\phi}_R(\bar{N}) = \begin{cases} \phi_R(\bar{N}) & \text{if } N_2 < L_2, \\ 2^{\sum_{k \neq 2} N_k} \phi_R(\bar{N}) & \text{if } N_2 \geq L_2. \end{cases} \quad (3.18)$$

Then $\hat{\mu}_i(\bar{n})$ can be written

$$\hat{\mu}_i(\bar{n}) = \mu_i \frac{\hat{\phi}_R(\bar{N} - E_i)}{\hat{\phi}_R(\bar{N})} \prod_{r \in R} \frac{\phi^{(r)}(\bar{n}^{(r)} - e_i^{(r)})}{\phi^{(r)}(\bar{n}^{(r)})}, \quad i \in C_s.$$

Note that the local service functions are not affected by this modification. As the service speed reduction is a global effect, this is what one should expect. Further note that in the case of a closed queueing network service speed reduction at all clusters $k \neq 2$ is equivalent to increasing the service speed at cluster 2 as can be seen from (3.18) by noting that $\sum_{k \neq 2} N_k = N - N_2$, where N is the total number of jobs present at the system. In the case of an open queueing network this equivalence is lost. $\square$

Example 3.15 (Stop protocol) In the previous example service at clusters $k \neq 2$ is slowed down if $N_2 \geq L_2$. A more drastic approach is to stop service at all clusters as well as arrivals from the outside if the number of jobs at cluster 2 reaches some upper bound U_2 . The service rate given in (3.17) is then modified to

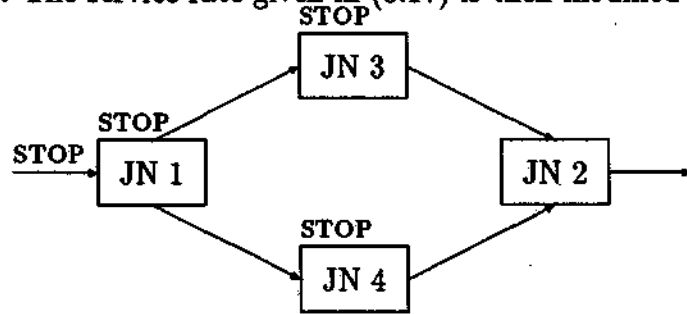


Figure 7. Stop service and arrivals

$$\tilde{\mu}_i(\bar{n}) = \begin{cases} \mu_i f_i(n_i) F_r(N_r), & i \in C_r, r = 2, \\ \mu_i f_i(n_i) F_r(N_r), & i \in C_r, r \neq 2, N_2 < U_2, \\ \mu_0, & i = 0, N_2 < U_2, \\ 0 & \text{otherwise.} \end{cases}$$

This service rate can be written in the form (2.1). To this end, with ϕ_R given in (3.16), define

$$\tilde{\psi}_R(\bar{N}) = \begin{cases} \phi_R(\bar{N}), & \text{if } N_2 < U_2, \\ 0, & \text{if } N_2 = U_2. \end{cases}$$

Then $\tilde{\mu}_i(\bar{n})$ can be written

$$\tilde{\mu}_i(\bar{n}) = \mu_i \frac{\hat{\psi}_R(\bar{N} - E_s)}{\phi_R(\bar{N})} \prod_{r \in R} \frac{\phi^{(r)}(\bar{n}^{(r)} - e_i^{(r)})}{\phi^{(r)}(\bar{n}^{(r)})}, \quad i \in C_s, \quad s \in R \cup \{0\}.$$

Note that this example cannot be obtained as a special limiting case of the previous example. In the first place, in the formulation of Example 3.14 we require positive service speeds. Secondly, the arrival rate μ_0 is not affected in Example 3.14, whereas here it has to be set equal to zero. $\square$

4 Norton's theorem

In Section 3 the transition rates (2.2) are decomposed into a global and local part such that the equilibrium distribution decomposes similarly. In Section 3.1, motivated by electrical circuit theory, the routing probabilities are decomposed such that the routing characteristics at global level and at local level can be analyzed separately. In Section 3.3, the service rates are decomposed such that the equilibrium distribution decomposes into a global and a local part. Although the form (3.14) is build up by separate components associated with global and local characteristics, it is not yet clear whether it can indeed be seen as a full decomposition of the equilibrium distribution into a purely global process and the marginal distributions of purely local processes, or allows us to aggregate subnetworks into single stations with appropriate service rates without affecting the equilibrium distribution of the rest of the queueing network. Based on the decomposition of the transition rates given in Section 3, this section derives these results. This section first gives Norton's theorem in weak form. This theorem allows us to separately calculate the global and local characteristics of the queueing network. Second, we give Norton's theorem in strong form. In addition to the applications of Norton's theorem in weak form, this theorem gives a formal justification for determining global characteristics via monitoring the behaviour of the queueing network at global level (e.g. monitoring the state dependent service rate and state dependent routing at global level) and shows that conditional on the global state the subnetworks are independent. Furthermore, from these theorema we obtain that aggregation of subnetworks into single queues does not affect the behaviour of the rest of the queueing network. This extends Norton's theorem to queueing networks with state dependent routing and servicing.

The following theorem establishes a decomposition of the process with transition rates (2.2) into a global process and a local process such that, except for normalization, the equilibrium distribution of the process decomposes into a global equilibrium distribution determined by the global process and a local equilibrium distribution factorizing into the subnetworks determined by the local processes at the subnetworks.

Theorem 4.1 (Norton's theorem in weak form)

(i) Service rates

If the process with transition rates (2.2) satisfies Assumptions 3.1, 3.11, 3.12 and Conditions 3.4, 3.5 then the service function (2.1) can be written as

$$\mu_i(\bar{n}) = M_r(\bar{N}) \prod_{s \in R} \mu_i^{(s)}(\bar{n}^{(s)}), \quad i \in C_r, \quad r \in R \cup \{0\}, \quad (4.1)$$

where M_r is the global service rate defined by

$$M_r(\bar{N}) = \begin{cases} \frac{\psi_R(\bar{N} - \bar{E}_r)}{\phi_R(\bar{N})}, & \text{if } r \in R, \\ \mu_0 \frac{\psi_R(\bar{N})}{\phi_R(\bar{N})}, & \text{if } r = 0, \end{cases} \quad (4.2)$$

while $\mu_i^{(s)}$ the local service rate at cluster s defined by

$$\mu_i^{(s)}(\bar{n}^{(s)}) = \begin{cases} \mu_i \frac{\psi^{(s)}(\bar{n}^{(s)} - e_i^{(s)})}{\phi^{(s)}(\bar{n}^{(s)})}, & \text{if } i \in C_s, \\ \frac{\psi^{(s)}(\bar{n}^{(s)})}{\phi^{(s)}(\bar{n}^{(s)})}, & \text{if } i \in C_r, \quad r \neq s. \end{cases} \quad (4.3)$$

(ii) Global and local processes

With B_R a normalizing constant, the distribution

$$\pi_R(\bar{N}) = B_R \phi_R(\bar{N}) H_R(\bar{N}) \quad (4.4)$$

is the equilibrium distribution of the global process with transition rates defined as

$$q_R(\bar{N}, \bar{N} - \bar{E}_r + \bar{E}_s) = M_r(\bar{N}) p^{rs}(\bar{N}), \quad r, s \in R \cup \{0\}. \quad (4.5)$$

With $B^{(r)}$ a normalizing constant, the distribution

$$\pi^{(r)}(\bar{n}^{(r)}) = B^{(r)} \phi^{(r)}(\bar{n}^{(r)}) H^{(r)}(\bar{n}^{(r)}) \prod_{k \in C_r} \left(\frac{1}{\mu_k} \right)^{n_k^{(r)}} \quad (4.6)$$

is the equilibrium distribution of the local process with transition rates defined as

$$q^{(r)}(\bar{n}^{(r)}, \bar{n}^{(r)} - e_i^{(r)} + e_j^{(r)}) = \mu_i^{(r)}(\bar{n}^{(r)}) p_{ij}^{(r)}(\bar{n}^{(r)}), \quad i, j \in C_r \cup \{0\}. \quad (4.7)$$

Furthermore, with B a normalizing constant,

$$\pi(\bar{n}) = B \pi_R(\bar{N}) \prod_{r \in R} \pi^{(r)}(\bar{n}^{(r)}), \quad \bar{n} \in S. \quad (4.8)$$

Proof (4.1) follows by insertion of Assumptions 3.11, 3.12 into (2.1). Insertion of (4.4), (4.5) and (4.6), (4.7) into the appropriate global balance equations establishes that (4.4) and (4.6) are indeed the equilibrium distributions at global and local level. Finally, (4.8) is a direct consequence of Theorem 3.6. $\square$

Remark 4.2 (Weak form) Although Theorem 4.1 above establishes a full decomposition of the equilibrium distribution, π , into a global component, π_R , and local components, $\pi^{(r)}$, $r \in R$, it does not enable us to compute the global part simply by monitoring the output processes of subnetworks. This can be seen by observing that π_R is not the equilibrium distribution of the aggregated network. Summing π over all states $\bar{n}$ such that $\sum_{j \in C_r} n_j = N_r$, $r \in R$, leads to the aggregated equilibrium distribution, Π_R :

$$\Pi_R(\bar{N}) \equiv \sum_{\bar{n}^{(r)}: N_r, r \in R} \pi(\bar{n}) = B\pi_R(\bar{N}) \prod_{r \in R} \pi^{(r)}(N_r), \quad (4.9)$$

where $\bar{n}^{(r)} : N_r$ is short for $\bar{n}^{(r)} : \sum_{j \in C_r} n_j = N_r$. As Theorem 4.1 establishes a full decomposition similar to the decomposition in Norton-type results it is called Norton's theorem in weak form. $\square$

Remark 4.3 (Implications of Norton's theorem in weak form) Norton's theorem in weak form establishes a decomposition of the process with transition rates (2.2) into a global process and a local process. This decomposition is such that:

- the global and local transition rates can be immediately obtained from the transition rates of the original process;
- the local process factorizes into the subnetworks;
- the global process and the local processes at the subnetworks can be analyzed separately;
- except for normalization, the equilibrium distribution of the original process consists only of a global part determined by the global process and a local part factorizing into the subnetworks determined by the local processes at the subnetworks. $\square$

As is discussed in Remark 4.2, in Norton's theorem we want a full decomposition of the equilibrium distribution into a global part and a local part such that the global part can be computed via monitoring the output processes of the subnetworks. To this end, the transition rates of the global process must be such that Π_R as defined in (4.9) is the equilibrium distribution of the global process and such that the probability flows between subnetworks in the global process equals the probability flows between subnetworks in the original non-aggregated process. The following theorem establishes this decomposition.

Theorem 4.4 (Norton's theorem in strong form)

(i) Global process

Define the global process with transition rates

$$Q(\bar{N}, \bar{N} - \bar{E}_r + \bar{E}_s) = \begin{cases} \frac{\Psi_R(\bar{N} - \bar{E}_r)}{\Phi_R(\bar{N})} p^{rs}(\bar{N}), & \text{if } r \in R, \\ \mu_0 \frac{\Psi_R(\bar{N})}{\Phi_R(\bar{N})} p^{0s}(\bar{N}), & \text{if } r = 0, \end{cases} \quad (4.10)$$

and service functions Ψ_R, Φ_R

$$\Psi_R(\bar{N}) = \psi_R(\bar{N}) \prod_{r \in R} \sum_{\bar{n}^{(r)}: N_r} \psi^{(r)}(\bar{n}^{(r)}) H^{(r)}(\bar{n}^{(r)}) \prod_{j \in C_r} \left(\frac{1}{\mu_j} \right)^{n_j^{(r)}}, \quad (4.11)$$

$$\Phi_R(\bar{N}) = \phi_R(\bar{N}) \prod_{r \in R} \sum_{\bar{n}^{(r)}: N_r} \phi^{(r)}(\bar{n}^{(r)}) H^{(r)}(\bar{n}^{(r)}) \prod_{j \in C_r} \left(\frac{1}{\mu_j} \right)^{n_j^{(r)}}. \quad (4.12)$$

If the process with transition rates (2.2) satisfies Assumptions 3.1, 3.11, 3.12 and Conditions 3.4, 3.5 then, with B_R a normalizing constant, the global process possesses an equilibrium distribution, Π_R , given by

$$\Pi_R(\bar{N}) = B_R \Phi_R(\bar{N}) H_R(\bar{N}). \quad (4.13)$$

(ii) Aggregated process

If for all $r \in R$ the service function $\psi^{(r)}$ satisfies the following technical assumption for all i and all N_r

$$\begin{aligned} \sum_{\bar{n}^{(r)}: N_r} \sum_{i \in C_r} \psi^{(r)}(\bar{n}^{(r)} - e_i^{(r)}) H^{(r)}(\bar{n}^{(r)}) \mu_i p_{i0}^{(r)}(\bar{n}^{(r)}) \prod_{j \in C_r} \left(\frac{1}{\mu_j} \right)^{n_j^{(r)}} \\ = \sum_{\bar{n}^{(r)}: N_r - 1} \psi^{(r)}(\bar{n}^{(r)}) H^{(r)}(\bar{n}^{(r)}) \prod_{j \in C_r} \left(\frac{1}{\mu_j} \right)^{n_j^{(r)}} \end{aligned} \quad (4.14)$$

then the global process equals the aggregated process, that is Π_R is the aggregated equilibrium distribution:

$$\Pi_R(\bar{N}) = \sum_{\bar{n}^{(r)}: N_r, r \in R} \pi(\bar{n}). \quad (4.15)$$

and the probability flow between aggregated subnetworks equals the probability flow between subnetworks in the original non-aggregated queueing network:

$$\Pi_R(\bar{N}) Q(\bar{N}, \bar{N} - \bar{e}_r + \bar{e}_s) = \sum_{\bar{n}: N_t, t \in R} \sum_{i \in C_r, j \in C_s} \pi(\bar{n}) q(\bar{n}, \bar{n} - e_i + e_j), \quad (4.16)$$

where for $r = s$ the summation in the right-hand side is over all transitions in which a job first leaves cluster r and then reroutes to cluster r .

(iii) Equilibrium distribution factorizes

With $\pi^{(r)}$ the equilibrium distribution of the local process at cluster r , and $\pi^{(r)}(\bar{n}^{(r)} | N_r)$ the conditional probability of $\bar{n}^{(r)}$ given N_r jobs at cluster r , the equilibrium distribution, π , of the original process can be written

$$\pi(\bar{n}) = \Pi_R(\bar{N}) \prod_{r \in R} \pi^{(r)}(\bar{n}^{(r)} | N_r). \quad (4.17)$$

Proof From Condition 3.4, insertion of Π_R as given in (4.13) into the appropriate global balance equations gives that Π_R is the equilibrium distribution of the global process.

Insertion of (4.4),(4.6),(4.8) into the right-hand side of (4.15) gives

$$\begin{aligned} \sum_{\bar{n}^{(r)}: N_r, r \in R} \pi(\bar{n}) &\stackrel{(4.8)}{=} \sum_{\bar{n}^{(r)}: N_r, r \in R} B \pi_R(\bar{N}) \prod_{r \in R} \pi^{(r)}(\bar{n}^{(r)}) \\ &\stackrel{(4.4),(4.6)}{=} \left(B B_R \prod_{r \in R} B^{(r)} \right) \Phi_R(\bar{N}) H_R(\bar{N}) \end{aligned}$$

which gives (4.15).

Insertion of the specific form (4.10) for Q , (4.13) for Π_R , (2.1),(2.2) for q , (2.5) for π , Assumptions 3.1, 3.11, 3.12, Conditions 3.4, 3.5 and Theorem 3.6 for H into (4.16) gives that (4.16) holds if and only if the following relation holds true.

$$\begin{aligned} B_R H_R(\bar{N}) \Psi_R(\bar{N} - \bar{E}_r) p^{rs}(\bar{N}) &= B H_R(\bar{N}) \psi_R(\bar{N} - \bar{E}_r) p^{rs}(\bar{N}) \\ &\times \sum_{\bar{n}: N_r, t \in R} \left(\prod_{k \neq r} \psi^{(k)}(\bar{n}^{(k)}) H^{(k)}(\bar{n}^{(k)}) \prod_{j \in C_k} \left(\frac{1}{\mu_j} \right)^{n_j^{(k)}} \right) \\ &\times \sum_{i \in C_r} \mu_i \psi^{(r)}(\bar{n}^{(r)} - e_i^{(r)}) H^{(r)}(\bar{n}^{(r)}) p_{i0}^{(r)}(\bar{n}^{(r)}) \prod_{j \in C_r} \left(\frac{1}{\mu_j} \right)^{n_j^{(r)}} \end{aligned}$$

From (4.11),(4.14) we obtain that this equation is satisfied.

Finally, from Theorem 4.1, if $\sum_{j \in C_r} n_j^{(r)} = N_r$

$$\pi^{(r)}(\bar{n}^{(r)} | N_r) = \frac{\phi^{(r)}(\bar{n}^{(r)}) H^{(r)}(\bar{n}^{(r)}) \prod_{j \in C_r} \left(\frac{1}{\mu_j} \right)^{n_j^{(r)}}}{\sum_{\bar{n}^{(r)}: N_r} \phi^{(r)}(\bar{n}^{(r)}) H^{(r)}(\bar{n}^{(r)}) \prod_{j \in C_r} \left(\frac{1}{\mu_j} \right)^{n_j^{(r)}}}.$$

From (4.12),(4.13) and Theorem 3.6 we now obtain (4.17). $\square$

Remark 4.5 (Comparison of Norton's theorem) At first glance, the results of Norton's theorem in strong form are similar to the results of Norton's theorem in weak form. However, the global process with transition rates q_R as defined in Theorem 4.1 is not the same as the global process Q defined in Theorem 4.4. In contrast to the global process q_R which can immediately be obtained from the transition rates, the global process Q must be computed from the transition rates and can not as easily be obtained as q_R . The result of Norton's theorem in strong form, however, is much stronger than the result of Norton's theorem in weak form. In Theorem 4.4 the decomposition into global and local distributions, (4.17), is complete whereas in (4.8) the normalizing constant relates the global and local processes and must still be computed. Furthermore, the global process of Theorem 4.4 describes the service rates of the subnetworks, whereas the global process of Theorem 4.1 merely describes the global part of the transition rates. $\square$

Remark 4.6 (Implications of Norton's theorem in strong form) Norton's theorem in strong form establishes a decomposition of the process with transition rates (2.2) into a global process and a local process such that:

- the local transition rates can be immediately obtained from the transition rates of the original process;
- the local process factorizes into the subnetworks;
- the global process is the process aggregated over the clusters, i.e. the global equilibrium distribution equals the aggregated equilibrium distribution and the probability flow between queues of the global process equals the probability flow between clusters of the original process;
- when the transition rates of the global process are determined, the global and local processes can be analyzed separately;
- the equilibrium distribution of the original process consists only of a global part determined by the global process and a local part factorizing into the subnetworks determined by the local processes at the subnetworks. $\square$

In addition to the points mentioned in Remark 4.6 above, the following corollary gives a major non-trivial result following from Theorem 4.4.

Corollary 4.7 (Conditional independence of subnetworks) *Conditional on the global state of the queueing network the subnetworks are independent:*

$$\pi(\bar{n}|\bar{N}) = \prod_{r \in R} \pi^{(r)}(\bar{n}^{(r)}|N_r), \quad \bar{n} \in S.$$

Proof From (4.15) and the definition of conditional probability we obtain that $\pi(\bar{n}) = \Pi_R(\bar{N})\pi(\bar{n}|\bar{N})$ and (4.17) completes the proof. $\square$

Remark 4.8 ($\psi = \phi$) If $\psi = \phi$ the service rate $\mu_i(\bar{n})$ as given in (4.1) in Theorem 4.1 depends on the global state and the local state of cluster r only:

$$\mu_i(\bar{n}) = M_r(\bar{N})\mu_i^{(r)}(\bar{n}^{(r)}), \quad i \in C_r.$$

Furthermore, note that the property $\psi = \phi$ carries over to the global process of Theorem 4.4, i.e. $\psi = \phi$ implies $\Psi_R = \Phi_R$, as can immediately be seen from (4.11),(4.12). $\square$

The main result of Norton's theorem as appearing in the literature is that the marginal distribution of the subnetwork of interest is the same in the original queueing network and in the partly aggregated queueing network. As the equilibrium distribution π as given in (2.5) can, due to the aggregation method, be written in the form (4.17) this result is an immediate consequence of Norton's theorem. This is stated in the following corollary, where for simplicity of formulation we have assumed that $\psi = \phi$.

Corollary 4.9 (Marginal distribution) Assume that $\psi = \phi$. Then the marginal distribution, $P(\bar{n}^{(k)})$, of $\bar{n}^{(k)}$ jobs at cluster k in the queueing network in which all clusters except cluster k are aggregated into single queues with service rates

$$M_r(\bar{N}) = \begin{cases} \frac{\Phi_R(\bar{N} - \bar{E}_r)}{\Phi_R(\bar{N})}, & \text{if } r \in R, r \neq k, \\ \mu_0, & \text{if } r = 0, \end{cases}$$

is the same as the marginal distribution, $\pi(\bar{n}^{(k)})$, of $\bar{n}^{(k)}$ jobs at cluster k in the original queueing network.

Proof From (4.1) and $\psi = \phi$ we obtain that the service rate $\mu_i(\bar{n})$ for servicing a job at station i of cluster k for the partly aggregated queueing network is given by

$$\mu_i(\bar{n}) = \frac{\phi_R(\bar{N} - \bar{E}_r)}{\phi_R(\bar{N})} \mu_i \frac{\phi^{(k)}(\bar{n}^{(k)} - e_i^{(k)})}{\phi^{(k)}(\bar{n}^{(k)})}, \quad i \in C_k.$$

From (4.12) this can be written

$$\mu_i(\bar{n}) = \frac{\Phi_R(\bar{N} - \bar{E}_k)}{\Phi_R(\bar{N})} \mu_i \frac{\frac{\phi^{(k)}(\bar{n}^{(k)} - e_i^{(k)})}{\sum_{\bar{n}^{(k)}: N_k - 1} \phi^{(k)}(\bar{n}^{(k)}) H^{(k)}(\bar{n}^{(k)}) \prod_{j \in C_k} \left(\frac{1}{\mu_j}\right)^{n_j^{(k)}}}}{\frac{\phi^{(k)}(\bar{n}^{(k)})}{\sum_{\bar{n}^{(k)}: N_k} \phi^{(k)}(\bar{n}^{(k)}) H^{(k)}(\bar{n}^{(k)}) \prod_{j \in C_k} \left(\frac{1}{\mu_j}\right)^{n_j^{(k)}}}},$$

where $H^{(k)}$ is the solution to the traffic equations for cluster k if cluster k would be considered in isolation. From Norton's theorem in weak form we now obtain

$$P(\bar{n}^{(k)}) \propto \sum_{\bar{N}: N_k = \sum_{j \in C_k} n_j^{(k)}} \Phi_R(\bar{N}) H_R(\bar{N}) \frac{\phi^{(k)}(\bar{n}^{(k)}) H^{(k)}(\bar{n}^{(k)}) \prod_{j \in C_k} \left(\frac{1}{\mu_j}\right)^{n_j^{(k)}}}{\sum_{\bar{n}^{(k)}: N_k} \phi^{(k)}(\bar{n}^{(k)}) H^{(k)}(\bar{n}^{(k)}) \prod_{j \in C_k} \left(\frac{1}{\mu_j}\right)^{n_j^{(k)}}}.$$

From Norton's theorem in strong form, by summation we obtain

$$\begin{aligned} \pi(\bar{n}^{(k)}) &= \sum_{\bar{n}: n_j = n_j^{(k)}} \pi(\bar{n}) \\ &= \sum_{\bar{N}: N_k = \sum_{j \in C_k} n_j^{(k)}} \Phi_R(\bar{N}) \pi^{(r)}(\bar{n}^{(k)} | N_k) \\ &= B_R \sum_{\bar{N}: N_k = \sum_{j \in C_k} n_j^{(k)}} \Phi_R(\bar{N}) H_R(\bar{N}) \frac{\phi^{(k)}(\bar{n}^{(k)}) H^{(k)}(\bar{n}^{(k)}) \prod_{j \in C_k} \left(\frac{1}{\mu_j}\right)^{n_j^{(k)}}}{\sum_{\bar{n}^{(k)}: N_k} \phi^{(k)}(\bar{n}^{(k)}) H^{(k)}(\bar{n}^{(k)}) \prod_{j \in C_k} \left(\frac{1}{\mu_j}\right)^{n_j^{(k)}}} \end{aligned}$$

which completes the proof. $\square$

Remark 4.10 (Assumptions justifying (4.14)) As is mentioned in Theorem 4.4, relation (4.14) is a technical assumption. This assumption is used only to show that (4.16) holds. Note that this technical assumption is not a very stringent condition as it is satisfied in many practical cases. For example, if the state space is such that the set of admissible states for subnetwork r is coordinate convex:

$$C^{(r)} = \{\bar{n}^{(r)} \mid \sum_{j \in C_r} n_j^{(r)} \leq U^{(r)}\}, \quad r \in R, \quad (4.18)$$

the left-hand side of (4.14) can be written

$$\begin{aligned} B^{(r)} \cdot \text{LHS} &= \sum_{\bar{n}^{(r)}: N_r} \sum_{i \in C_r} \pi^{(r)}(\bar{n}^{(r)}) q^{(r)}(\bar{n}^{(r)}, \bar{n}^{(r)} - e_i^{(r)}) \\ &= \sum_{\bar{n}^{(r)}: N_r} \sum_{i \in C_r} \sum_{j \in C_r \cup \{0\}} \pi^{(r)}(\bar{n}^{(r)}) q^{(r)}(\bar{n}^{(r)}, \bar{n}^{(r)} - e_i^{(r)} + e_j^{(r)}) \\ &\quad - \sum_{\bar{n}^{(r)}: N_r} \sum_{i \in C_r} \sum_{j \in C_r} \pi^{(r)}(\bar{n}^{(r)}) q^{(r)}(\bar{n}^{(r)}, \bar{n}^{(r)} - e_i^{(r)} + e_j^{(r)}) \\ &\stackrel{(3.3)}{=} \sum_{\bar{n}^{(r)}: N_r} \sum_{i \in C_r} \sum_{j \in C_r \cup \{0\}} \pi^{(r)}(\bar{n}^{(r)} - e_i^{(r)} + e_j^{(r)}) q^{(r)}(\bar{n}^{(r)} - e_i^{(r)} + e_j^{(r)}, \bar{n}^{(r)}) \\ &\quad - \sum_{\bar{n}^{(r)}: N_r} \sum_{i \in C_r} \sum_{j \in C_r} \pi^{(r)}(\bar{n}^{(r)}) q^{(r)}(\bar{n}^{(r)}, \bar{n}^{(r)} - e_i^{(r)} + e_j^{(r)}) \\ &= \sum_{\bar{n}^{(r)}: N_r} \sum_{i \in C_r} \pi^{(r)}(\bar{n}^{(r)} - e_i^{(r)}) q^{(r)}(\bar{n}^{(r)} - e_i^{(r)}, \bar{n}^{(r)}) \\ &\stackrel{(4.18)}{=} \sum_{\bar{n}^{(r)}: N_r - 1} \sum_{i \in C_r} \pi^{(r)}(\bar{n}^{(r)}) q^{(r)}(\bar{n}^{(r)}, \bar{n}^{(r)} + e_i^{(r)}), \end{aligned}$$

which immediately implies that (4.14) is satisfied. $\square$

Conclusions 4.11 (Practical applications of Norton's theorem) Norton's theorem establish a decomposition of the process with transition rates (2.2) into a global process and a local process. As is discussed in Remark 4.6, the decomposition into a global process and a local process is more difficult to establish in Norton's theorem in strong form than in Norton's theorem in weak form. The result of Norton's theorem in strong form, however, is indeed stronger than the result of Norton's theorem in weak form.

The following points characterize practical applications of Norton's theorem. The first two points are valid for both theorems. The last three points are valid for Norton's theorem in strong form only.

- (1) For large queueing networks, determination of a solution to the (state dependent) traffic equations is (computationally) very hard. Norton's theorem allows us to analyze the global and local traffic equations separately. As the local traffic equations factorize into the subnetworks, this gives much smaller problems. The solution to the total traffic equations is just the product of the global and the local parts.
- (2) Calculation of the normalization constant for large processes (i.e. processes with a large state space) is, in general, very difficult. Norton's theorem allows us to first analyze the normalization constants at local level and subsequently to analyze the normalization constant at global level. As the local and global state spaces are substantially smaller than the total state space, these problems are substantially smaller than the original problem.
- (3) If the global characteristics of the queueing network, i.e. characteristics depending on the global state of the queueing network only, are of interest, Norton's theorem in strong form allows us to only analyze the queueing network at global level. To this end, note that the global routing probabilities can be determined by observing the routing between clusters in the original queueing network. The global service functions, Ψ_R , Φ_R , can be obtained by monitoring the state dependent speed at which jobs leave the clusters. From the thus obtained transition rates the global equilibrium distribution can be obtained. From this global equilibrium distribution the global characteristics can be determined.
- (4) If characteristics depending on the local state of a single cluster are of interest only, Norton's theorem in strong form allows us to analyze the queueing network at a global level as discussed in (3). For the cluster of interest, Norton's theorem allows us to analyze this cluster with transition rates as given in Theorem 4.1 separately. The equilibrium distribution of the queueing network containing full information on the state of the cluster of interest is then given by the global equilibrium distribution multiplied with the distribution of the cluster at interest conditional on the global state of that cluster. This method can also be applied when characteristics depending on the local state of several clusters are of interest.
- (5) If the transition rates of the clusters in isolation (i.e. not part of the queueing network) are known, then the service functions Ψ_R , Φ_R can simply be calculated from the functions $\psi^{(r)}$ and $\phi^{(r)}$ as obtained from the transition rates of the clusters in isolation. When Ψ_R and Φ_R are calculated, the global behaviour of the queueing network can be analyzed without considering the local behaviour.

5 Examples

5.1 Global throughput determines local behaviour; workload balancing

This example shows that under global blocking conditions such as capacity constraints on the total number of jobs at the clusters it is possible to obtain the local equilibrium distribution from the global throughput of the clusters and thus to obtain both global and local performance measures by merely analyzing the global throughputs. To this end, consider a queueing network consisting of N stations grouped into R clusters such that the routing probabilities have the form given in Assumption 3.1. Assume that we want the workload to be balanced over the clusters in the sense that the total number of jobs at the clusters is constrained to be in the range $L_r \leq N_r \leq U_r$, $r \in R$. Furthermore, except for effects due to the boundaries, assume that the service rate at a cluster only depends on the state of that cluster and that the global routing probabilities are state independent except for blocking phenomena arising at the boundaries. As all blocking effects are due to global requirements on the workload at the clusters the local transition rates are not affected by these blocking effects, i.e. all blocking effects are incorporated in the global process. Let p^{rs} be the state independent part of the global routing probabilities and $\{a^{(r)}\}_{r \in R}$ the solution to the global traffic equations (3.7) as given in Example 3.8. Similar to Example 3.9, for Condition 3.4 to be satisfied, the global routing probabilities must be modified according to a blocking protocol. The following blocking protocol is used:

- (1) Jobs leaving cluster r when $N_r = L_r$ are rerouted into cluster r . This guarantees the minimal workload.
- (2) Jobs leaving cluster $s \neq r$, when $N_r = U_r$ are rerouted into cluster s according to the global rerouting protocol. This guarantees that the upper boundaries are not exceeded.

By noting that the set of admissible states is restricted to

$$C = \{\bar{N} | L_r \leq N_r \leq U_r, \text{ for all } r \in R, \text{ and } N_r + N_s < U_r + U_s \text{ for any } r \neq s\},$$

the global traffic equations (3.2) now read for $r \in R$, $\bar{N} \in C$:

$$\begin{aligned} H_R(\bar{N}) = & \sum_{s \in R, s \neq r} H_R(\bar{N} - \bar{E}_r + \bar{E}_s) \mathbf{1}(\bar{N} - \bar{E}_r + \bar{E}_s \in C) \mathbf{1}(N_k < U_k \text{ for all } k \neq s) p^{sr} \\ & + H_R(\bar{N}) \mathbf{1}(N_k = U_k \text{ for some } k \neq r \text{ or } N_r = L_r). \end{aligned}$$

As $\bar{N} - \bar{E}_r + \bar{E}_s \notin C$ if $N_k = U_k$ for $k \neq r$ or if $N_r = L_r$, we obtain that with

$$H_R(\bar{N}) = \prod_{r \in R} (a^{(r)})^{N_r}, \quad \bar{N} \in C, \quad (5.1)$$

Condition 3.4 is satisfied. This specifies the global behaviour of the queueing network.

Now let us consider the local behaviour. Up to normalization constants, $a^{(r)}$ equals the global throughput of cluster r , i.e. $a^{(r)}$ equals the parameter of the arrival process to cluster r . The transition rates for the local process at cluster r with unit arrival parameter are given in (4.7). From Condition 3.5 we now obtain that the local equilibrium distribution at cluster r is given in (4.6), where $H^{(r)}$ is obtained by analyzing cluster r with unit arrival parameter. From Remark 2.6 we now obtain that the local equilibrium distribution of cluster r with input rate $a^{(r)}$ is given by

$$\pi^{(r)}(\bar{n}^{(r)}) = B^{(r)} \phi^{(r)}(\bar{n}^{(r)}) H^{(r)}(\bar{n}^{(r)}) \prod_{j \in C_r} \left(\frac{a^{(r)}}{\mu_j} \right)^{n_j^{(r)}}. \quad (5.2)$$

From Norton's theorem in weak form 4.1 we immediately obtain that the marginal distribution of cluster r , $\pi(\bar{n}^{(r)})$, is then given by

$$\pi(\bar{n}^{(r)}) = \pi^{(r)}(\bar{n}^{(r)}). \quad (5.3)$$

Note that the global blocking protocol used here is not essential to the derivation of (5.3). For example, an alternative blocking protocol is the stop protocol, where servicing is stopped at all non-saturated clusters if the number of jobs at a cluster reaches its upper bound. For example, in Example 3.9 this requires that the servicing at all stations 1, 2, 3 and 7, 8, 9 is stopped when cluster 2 is saturated, i.e. when $N_2 = U_2$. As is shown in [14], [15], the "stop protocol" and "recirculate protocol" are equivalent under product form conditions, that is under Assumptions 3.1, 3.11, 3.12 and Conditions 3.4, 3.5. Essential to the derivation of (5.2) is that H_R has the form (5.1) as this form allows us to obtain the throughput, $a^{(r)}$, of cluster r .

5.2 Internal blocking

In all examples given so far emphasis was on the global behaviour of the queueing network. Most particularly, global routing and blocking is treated in great detail. From the decomposition of the routing probabilities, however, the internal routing probabilities $p_{ij}^{(r)}$ can be state dependent too. From Assumption 3.1, as subnetwork we can insert all examples appearing in the literature of product form queueing networks with state dependent routing in which, except for a constraint on the total number of jobs present at the queueing network, arrivals cannot be blocked, i.e. for all $\bar{n}^{(r)}$ jobs arriving at the cluster are accepted at a station of the cluster:

$$\sum_{i \in C_r} p_{0i}^{(r)}(\bar{n}^{(r)}) = 1.$$

As an illustrative example, consider the queueing network consisting of two clusters as depicted in Figure 8, where cluster 1 is an arbitrary Jackson network and cluster 2 consists of three queues. Due to capacity constraints, the number of jobs at station 2 is constrained not to exceed Z_2 and the total number of jobs at cluster 2 is not to exceed U_2 . Assume that, except for blocking phenomena, the routing probabilities are as

indicated in Figure 8. If we use the recirculation protocol, if station 2 is saturated jobs leaving station 1 and 3 to route to station 2 are blocked and recirculated into station 1 and 3 respectively. If cluster 2 is saturated, jobs leaving cluster 1 are recirculated into cluster 1. This gives the following modification to the routing probabilities:

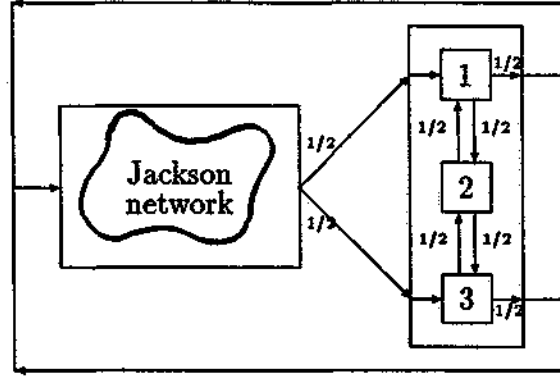


Figure 8. Internal blocking due to capacity constraints M_5

The local routing probabilities at cluster 2 are given by:

$$p_{21}^{(2)}(\bar{n}^{(2)}) = p_{23}^{(2)}(\bar{n}^{(2)}) = \frac{1}{2}, \quad \text{for all } \bar{n}^{(2)},$$

$$p_{10}^{(2)}(\bar{n}^{(2)}) = p_{30}^{(2)}(\bar{n}^{(2)}) = \frac{1}{2}, \quad \text{for all } \bar{n}^{(2)},$$

but

$$p_{12}^{(2)}(\bar{n}^{(2)}) = p_{32}^{(2)}(\bar{n}^{(2)}) = \frac{1}{2}, \quad \text{if } n_2 < Z_2,$$

$$p_{11}^{(2)}(\bar{n}^{(2)}) = p_{33}^{(2)}(\bar{n}^{(2)}) = \frac{1}{2}, \quad \text{if } n_2 = Z_2.$$

The global routing probabilities are given by:

$$p^{12}(\bar{N}) = 1, \quad \text{if } N_2 < U_2,$$

$$p^{11}(\bar{N}) = 1, \quad \text{if } N_2 = U_2.$$

The standard solution (3.9) still applies to both global and local level.

5.3 Nested aggregation

The aggregation procedure described in this paper can be nested. To describe this roughly, consider a queueing network consisting of N stations. If the stations can be grouped into clusters, say clusters $1, \dots, R$, such that the routing probabilities satisfy Assumption 3.1 and Conditions 3.4, 3.5 and the service rate satisfies Assumptions 3.11, 3.12, from Norton's theorem in strong form we obtain that the clusters can be aggregated into single stations. We then obtain a new queueing network consisting of R stations. At this new queueing network again we can apply the aggregation procedure to obtain a queueing network consisting of $R' < R$ stations. Nesting of aggregation allows us to first aggregate simple structures for which the aggregated service rates can easily be computed, such as for example a number of $M|M|\infty$ queues or a tandem line and subsequently aggregate the simplified clusters into single queues.

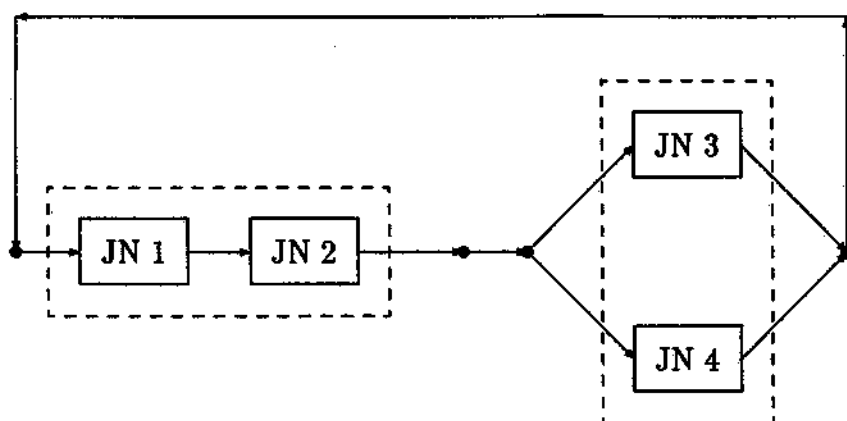


Figure 9. Nested aggregation

As the possibilities are non-exhaustive, while each additional nesting basically remains the same, let us merely give one simple example. Consider the queueing network as depicted in Figure 9. In the first aggregation step cluster i is aggregated into a single station, $i = 1, 2, 3, 4$. We thus obtain a 4 station queueing network. In this queueing network station 1 and station 2 can be aggregated into a single station, for example with a total capacity constraint, and stations 3 and 4 can be aggregated into a single station, for example with service delay depending on the total number of jobs at stations 3 and 4. We thus obtain a closed 2 station queueing network that can be analyzed by using standard methods.

6 Literature and concluding remarks

Remark 6.1 (Literature on Norton's theorem) Norton's theorem for queueing networks is introduced in [5] and states that the marginal distribution of a subnetwork, say σ , in the original queueing network equals the marginal distribution of σ when the rest of the queueing network is aggregated into a single queue with service rate set equal to the throughput of σ when the service rate of all stations in σ is reduced to zero. [5] prove the aggregation method to be correct for queueing networks of the BCMP-type in which σ consists of a single queue only. The proof given in this reference can easily be extended to the case where σ consists of several stations such that jobs enter σ through a single station and leave σ from a single station. In [1], [11] Norton's theorem is generalized to include BCMP-networks in which the routing between subnetworks is allowed to be arbitrary, but state-independent. Norton's theorem is further generalized in [17], where queueing networks consisting of two arbitrary quasireversible subnetworks are discussed. In all of these references the proof is based on the fact that the equilibrium distribution factorizes into the subnetworks and in addition that the routing between subnetworks is state independent. (As these references consider queueing networks consisting of two subnetworks only, the second observation states that jobs arriving at a subnetwork are always accepted.) Summarizing, from these references we obtain that:

- 1a subnetworks are analyzed as closed queueing networks;
- 2a the service rate of the aggregated queue cannot depend on the state of σ as the state of σ is deleted from the description when the throughput of σ with zero service rate is determined;
- 3a the global routing probabilities cannot be state dependent which excludes blocking.

In the present paper, the decomposition and aggregation of subnetworks is treated differently. We obtain that:

- 1b subnetworks are analyzed as open queueing networks;
- 2b the service rate of the aggregated queues is allowed to depend on the full global state;
- 3b global blocking is included as the global routing probabilities are state dependent.

Note that 2b and 3b immediately generalize 2a and 3a. Comparison of 1a and 1b shows that subnetworks are analyzed differently. This may cause different internal behaviour of the aggregated queues.

In contrast, note that the routing probabilities between subnetworks in [1], [11] are more general than our routing probabilities for the state independent case. In [1], [11] arbitrary routing probabilities $p_{ij}, i \in C_r, j \in C_s, r, s = 1, 2$, are included, whereas we assume $p_{ij} = p_{i0}^{(r)} p_{0j}^{(s)}$. This specific choice is due to the inclusion of blocking phenomena in the local routing probabilities. Note that arbitrary state independent routing probabilities can easily be incorporated into our framework. In this case the routing probabilities have the form $p_{ij}(\bar{n}) = p_{ij} p^{rs}(\bar{N})$. If a solution to the traffic equations for p_{ij} exists the local part of the solution to the traffic equations factorizes into the stations (cf. [14]). However, in this case the local process does not factorize into the subnetworks. As emphasis is on a complete factorization into the subnetworks of the local equilibrium distribution and process, which allows us to analyze subnetworks independently, the routing probabilities are chosen to be of the form (3.1). Finally, although not mentioned in [1], [5], [11], and [17], Norton's theorem as presented by these authors can be generalized to allow aggregation of the equilibrium distribution of several subnetworks independently. Again, state dependent routing cannot be incorporated in the generalization of Norton's theorem as presented in these references. $\square$

Remark 6.2 (Literature on factorization into subnetworks) Results similar to the factorization results obtained here have been reported in the literature. In [12] reversible subnetworks are connected via state independent global routing and in [10], [18] quasireversible subnetworks are connected via state dependent global routing. In these references, the transition rates depend on the local state only and a factorization of the equilibrium distribution similar to (4.8) for $B\pi_R = 1$ is obtained. As the present paper allows global state dependence also, it generalizes these results. $\square$

Remark 6.3 (Weak coupling) In [19, chapters 10, 11] the concept of weak coupling is described. A system is said to be weakly coupled if the equilibrium distribution consists of a part determined by the global routing probabilities and a part independent of the global routing. Thus, weak coupling is a property of the equilibrium distribution. In our notation both (4.8) and (4.17) state that the system is weakly coupled as the equilibrium distribution depends on the global routing characteristics through the global components, π_R and Π_R , only.

If the global routing is state independent, based on a condition similar to global balance for each cluster separately, [19] proves the system to be weakly coupled. If the global routing is state dependent, however, the condition stated in [19] for proving weak coupling is that a global balance-like condition is satisfied for the process with global routing removed from the routing characteristics, i.e. with $p^{rs}(\cdot)$ deleted from the routing probabilities $p_{ij}(\cdot)$. In the case of state independent routing [19] established a decomposition of the equilibrium distribution into components for each subnetwork. In the case of state dependent routing, however, such a decomposition cannot be concluded.

Also, the approach taken in [19] is different from the approach taken in the present paper. In [19] emphasis is on establishing a decomposition of the equilibrium distribution into a part determined by the global routing and a part independent of the global routing. A decomposition of the local part is not established. Furthermore, an essential part of Norton's theorem is that the process decomposes into a global process and a local process. This is guaranteed by our assumptions and cannot be concluded from the parameterization in [19]. Finally, in [19] global balance-like conditions are used when proving weak coupling. Although global balance is more general than local balance as guaranteed by the traffic equations, in queueing networks one usually requires that the traffic equations possess a solution. Therefore, our assumptions on the routing part of the transition rates are less general, but are sufficiently general in practical applications. $\square$

Remark 6.4 (Relation to electrical circuit theory) In Remark 6.1 above the queueing network analog of Norton's theorem as described in the literature is compared to the queueing network analog of Thevenin's theorem as described in the present paper. Although Norton's and Thevenin's theorem for electrical circuits are discussed extensively in the introduction, up to this moment Norton's and Thevenin's theorem for queueing networks have not been related to these theorems. Moreover, the distinction between Thevenin's theorem and Norton's theorem for queueing networks is not yet justified. To this end, reconsider Norton's and Thevenin's theorem for electrical circuits. From Figure 2 we obtain that the replacement circuit for Norton's theorem is determined by analyzing the current I flowing in a closed loop version of the electrical circuit. Similarly, from Figure 3 we obtain that the replacement circuit for Thevenin's theorem is determined by analyzing the potential difference V in an open version of the electrical circuit. Current flowing in the electrical circuit is the electrical analog of customers flowing in the queueing network (throughput). This justifies the name Norton's theorem for queueing networks as this method is based on analyzing the throughput in a closed loop version of the subnetwork. The potential

difference can be seen as the electrical analog of the service potential $\frac{\Psi_R}{\Phi_R}$ (cf. [19]). As our method is based on determining the service potential in an open network version of the subnetworks the name Thevenin's theorem for queueing networks would be justified. Note that, as Norton's theorem has become the standard phrasing in the literature on aggregation in queueing networks, we have called our results Norton's theorem. $\square$

References

- [1] Balsamo, S. and Iazeolla, G. (1982) "An extension of Norton's theorem for queueing networks." *IEEE Trans. on Software Eng.* **8**, 298-305.
- [2] Baskett, F. Chandy, K. M., Muntz, R. R. and Palacios, F. G. (1975) "Open, closed and mixed networks of queues with different classes of customers." *J. A. C. M.* **22**, 248-260.
- [3] Belevitch, V. (1968). *Classical network theory*. Holden-Day.
- [4] Boucherie, R. J. and Van Dijk, N. M. (1991). "Product forms for queueing networks with state dependent multiple job transitions." *Adv. Appl. Prob.* **23**, 152-187.
- [5] Chandy, K. M., Herzog, U. and Woo, L. (1975) "Parametric analysis of queueing networks." *IBM J. Res. Develop.* **19**, 36-42.
- [6] Conway, A. E. and Georganas, N. D. (1986) "Decomposition and aggregation by class in closed queueing networks." *IEEE Trans. on Software Eng.* **12**, 1025-1040.
- [7] Conway, A. E. (1990) "A new method of parametric analysis for product-form queueing networks." *Performance '90* (Eds. P. J. B. King, I. Mitrani and R. J. Pooley), North-Holland.
- [8] Courtois, P. J. (1977) *Decomposability: queueing and computer system applications*. Academic Press.
- [9] Henderson, W. and Taylor, P. G. (1990). "Networks of queues with batch arrivals and batch services." *Queueing Systems* **6**, 71-88.
- [10] Kelly, F. P. (1979). *Reversibility and stochastic networks*. Wiley.
- [11] Krizinger, P. S., Van Wyk, S. and Krzesinski, A. E. (1982) "A generalisation of Norton's theorem for multiclass queueing networks." *Performance Evaluation* **2**, 98-107.
- [12] Pollett, P. K. (1986) "Connecting reversible Markov processes." *Adv. Appl. Prob.* **18**, 880-900.

- [13] Serfozo, R. F. (1989). "Markovian network processes: Congestion-dependent routing and processing." *QUESTA* 5, 5-36.
- [14] Van Dijk, N. M.(1991). "Product forms for queueing networks with limited clusters." *Research Report, Free University, Amsterdam, The Netherlands*.
- [15] Van Dijk, N. M.(1991) "Stop=recirculate for exponential product form queueing networks with departure blocking. *Operations Research Letters* 10, 343-351.
- [16] Vantilborgh, H. (1978) "Exact aggregation in exponential queueing networks." *J. A. C. M.* 25, 620-629.
- [17] Walrand, J. (1983). "A note on Norton's theorem for queueing networks." *J. Appl. Prob.* 20, 442-444.
- [18] Walrand, J. and Varaiya, P. (1980) "Interconnections of Markov chains and quasi-reversible queueing networks." *Stoch. Proc. Appl.* 10, 209-219.
- [19] Whittle, P. (1986). *Systems in stochastic equilibrium*. Wiley.

1990-1	B. Vogelvang	Testing For Co-Integration with Spot Prices of Some Related Agricultural Commodities	1990-19	F.A.G. den Butter R.F. v.d. Wijngaert	Who is Correcting the Error ? A Co-integration Approach for Wages, Wage Space and Labour Conflicts in the Netherlands
1990-2	J.C.J.M. van den Bergh P. Nijkamp	Ecologically Sustainable Economic Development Concepts and Model Implications	1990-20	J.P. de Groot R. Ruben	Sistemas de Producción y Transferencia de Tecnología en la Economía Cafetalera de Centroamerica
1990-3	J.C.J.M. van den Bergh P. Nijkamp	Ecologically Sustainable Economic Development in a Regional System: A Case Study in Agricultural Development Planning in the Netherlands	1990-21	R. Ruben	Campeinado y Reforma Agraria en El Salvador
1990-4	C.Gorter P.Nijkamp P.Rietveld	Employers' Recruitment Behaviour and Re-Employment Probabilities of Unemployed	1990-22	J. van Ours G. Ridder	Vacancies and the Recruitment of New Employees
1990-5	K.Burger	Off-farm income and the farm-household the case of Kenyan smallholders	1990-23	A.F. de Vos J.J. de Vries	The Likelihood Function of a Generalized Gravity Model; Handling the Implicit Singularity of a Nonlinear Transformation
1990-6	H. Visser	Crowding out and the Government Budget	1990-24	D. van der Wal	Geloofwaardigheid en het Nederlandse wisselkoersbeleid
1990-7	P. Rietveld	Ordinal Data in Multicriteria Decision Making, a Stochastic Dominance Approach to Siting Nuclear Power Plants	1990-25	R.J. Veldwijk M. Boogaard M.V. van Dijk E.R.K. Spoor	EDSOs, implosion and explosion: concepts to automate a part of application maintenance
1990-8	G. van der Laan P.H.M. Ruys D.J.J. Taltman	Signaling devices for the supply of semi-public goods	1990-26	B. Hanzon	The area enclosed by the (oriented) Nyquist diagram and the Hilbert-Schmidt-Hankel norm of a linear system
1990-9	F.A.G. den Butter	Labour Productivity Slowdown and Technical Progress: An empirical analysis for The Netherlands	1990-27	R.W. van Zijp	Why Lucas is not a Hayekian
1990-10	R.W. van Zijp	Neo-Austrian Business Cycle Theory	1990-28	J. Rouwendal	On discrete choice under uncertainty: A generalization of the logit model and its application
1990-11	J.C. van Ours	Matching Unemployment and Vacancies: The Efficiency of the Dutch Labour Market	1990-29	J. Rouwendal	On the equitable distribution of the housing stock
1990-12	B. Vogelvang	Hypotheses Testing Concerning Relationships between Spot Prices of Various Types of Coffee	1990-30	J. Rouwendal	Stochastic market equilibria with efficient rationing (with an application to the Dutch housing market)
1990-13	A.F. de Vos I.J. Steyn	Stochastic Nonlinearity: A Firm Basis for the Flexible Functional Form	1990-31	J.A. Vijlbrief	The effects of unemployment insurance on the labour market
1990-14	Y.H. van Emmerik D. de Jong W.W.A. Zuurmond D.N. Dukkers-van Emden	Opereren in overleg: geprotocolleerde samenwerking 1e-2e-lijn bij dagchirurgie	1990-32	J.G.W. Simons H.P. Wansink	Traffic Ban, a means to combat smog?
1990-15	T.J.J.B. Wolters	Mediation and Collective Bargaining: A Diagnostic Approach	1990-33	J.C. van Ours T. Zoethout	De Interne Arbeidsmarkt van de Gemeente Amsterdam
1990-16	E.M.A. Scholten J. Koelwijjn	Financieringsproblematiek van startende ondernemingen: een mogelijke verklaring op basis van empirisch onderzoek.	1990-34	H.J. Bierens	A Note On the Limiting Distribution of Sample Autocorrelations in the Presence of a Unit Root
1990-17	E. Hüner H.P. Smit	Saturation and Model Specification of Passenger car Ownership	1990-35	T. Kuhlman	The Economic Integration of Refugees in Developing Countries: A Research Model.
1990-18	F.A.G. den Butter	Sociale zekerheid, de wig en economische groei	1990-36	T. Kuhlman	Towards a Definition of Refugees.
			1990-37	T. Kuhlman	Organized versus Spontaneous Settlement of Refugees in Africa.
			1990-38	R. Zuidema	The Neo-Austrian View on Interest
			1990-39	G.v.d.Laan	General Equilibrium in a Closed International Trade Model